EDITORIAL STAFF

EDITOR
WILLIAM KRUSKAL

ASSOCIATE EDITORS

ALLAN BIRNBAUM
Z. W. BIRNBAUM
DONALD A. DARLING
WASSILY HOEFFDING
N. L. JOHNSON
OSCAR KEMPTHORNE
E. L. LEHMANN

WITH THE COOPERATION OF

R. L. ANDERSON
R. C. BOSE
D. G. CHAPMAN
K. L. CHUNG
W. S. CONNOR
D. R. COX
J. F. DALY
CYRUS DERMAN
J. L. DOOB
MEYER DWASS
D. A. S. FRASER
SAMUEL KARLIN
HARRY KESTEN
C. H. KRAFT
SOLOMON KULLBACH
EUGENE LUKACS
F. J. MASSEY
G. E. NOETHER
HOWARD RAIFFA
H. E. ROBBINS
WALTER L. SMITH
D. R. TRUAX
LIONEL WEISS

PAST EDITORS OF THE ANNALS

H. C. CARVER, 1930–1938
S. S. WILKS, 1938–1949
T. W. ANDERSON, 1950–1952
E. L. LEHMANN, 1953–1955
T. E. HARRIS, 1955–1958

Published quarterly by the Institute of Mathematical Statistics in March, June, September and December at Baltimore, Maryland.

IMS INSTITUTIONAL MEMBERS

AMERICAN VISCOSE CORPORATION, Marcus Hook, Pennsylvania.
BELL TELEPHONE LABORATORIES, INC., TECHNICAL LIBRARY, 463 West Street, New York 14, New York.
COMMITTEE ON STATISTICS, INDIANA UNIVERSITY, Bloomington, Indiana
INTERNATIONAL BUSINESS MACHINES CORPORATION, New York
IOWA STATE COLLEGE, STATISTICAL LABORATORY, Ames, Iowa
LOCKHEED AIRCRAFT CORPORATION, Burbank, California
MASSACHUSETTS INSTITUTE OF TECHNOLOGY, HAYDEN LIBRARY, PERIODICAL DEPARTMENT, Cambridge 39, Massachusetts
MICHIGAN STATE UNIVERSITY, DEPARTMENT OF STATISTICS, East Lansing, Michigan
NATIONAL SECURITY AGENCY, Fort George G. Meade, Maryland
PRINCETON UNIVERSITY, DEPARTMENT OF MATHEMATICS, SECTION OF MATHEMATICAL STATISTICS, Princeton, New Jersey
PURDUE UNIVERSITY, Lafayette, Indiana
STATE UNIVERSITY OF IOWA, Iowa City, Iowa
THE CATHOLIC UNIVERSITY OF AMERICA, STATISTICAL LABORATORY, DEPARTMENT OF MATHEMATICS, Washington, D. C.
THE RAMO-WOOLDRIDGE CORPORATION, Los Angeles, California
UNIVERSITY OF CALIFORNIA, STATISTICAL LABORATORY, Berkeley, California
UNIVERSITY OF ILLINOIS, Urbana, Illinois
UNIVERSITY OF NORTH CAROLINA, DEPARTMENT OF STATISTICS, Chapel Hill, North Carolina
UNIVERSITY OF WASHINGTON, LABORATORY OF STATISTICAL, RESEARCH, Seattle, Washington

INVARIANCE THEORY AND A MODIFIED MINIMAX PRINCIPLE

By Oscar Wesler[1]

Stanford University and *University of Michigan*

TABLE OF CONTENTS

1. Introduction and summary 1
2. The use of invariance as a slicing principle 3
3. The principle of invariance in testing hypotheses 6
4. The modified minimax principle, completeness, and the generalized Hunt-Stein theorem 8
5. Mixed games, and the use of previous experience 18
Acknowledgment 19
References 20

1. Introduction and summary. One of the unpleasant facts about statistical decision problems is that they are generally much too big or too difficult to admit of practical solutions, a fact that is threatening to widen even further the gap between the theory and application of this brave new discipline. Briefly, the situation is this. For each possible decision procedure φ, the statistician is concerned only with the values $\rho(\omega, \varphi)$ of the risk function as ω ranges over the set Ω of all possible states of nature, so that a choice of a decision procedure amounts to a choice of a risk function. The obvious difficulty of comparing functions in the search for a best procedure now arises, constituting a major problem for the statistician. The Bayes and minimax principles, it should be noted, represent but two extremes to which the statistician can go to get around this difficulty rather than to meet it, the one assuming complete knowledge of an a priori probability distribution λ of the possible states ω, the other assuming least favorable circumstances about ω, so that in either case one considers only a single number per procedure rather than the entire function—the Bayesian the average risk with respect to λ, the minimax-man in all timidity the supremum of the risk function—comparisons thus becoming trivial in principle and obliging one to look simply for that procedure which minimizes these numbers. Inasmuch as the situations occurring in practice with regard to prior knowledge about ω usually lie between the two extremes just described, to this extent at least both principles are open to criticism. In view of the nature of the difficulty in choosing among procedures, the notion of admissible or complete classes of strategies is generally felt to provide the most satisfactory solution. Whereas it may be difficult to say what to do in a statistical decision problem, it is generally easier

Received March 7, 1958.

[1] This paper is the author's doctoral dissertation, accepted by Stanford University and published as Technical Report No. 35, Statistics Laboratory, Stanford University, October 6, 1955; research carried out under the sponsorship of the Office of Naval Research.

to say what not to do, so that the statistician separates out from consideration all inadmissible strategies and presents the practical man with what is at best a minimal essentially complete class of procedures. The choice of one from among these admissible procedures is then left to the best judgment, intuition, and past experience of the practical man. If the class is a small one, we have then achieved everything one can ask for, for the actual choice will then easily be made. The difficulty, however, is that the classes are usually much too big to be of real help.

The trouble, in a word, lies in the fact that it is the space Ω itself of possible states that is generally too big: one simply cannot look at and assess all the values of ω for each decision function φ. Now when problems turn out to be too big for practical purposes, it is natural to look for ways of cutting them down to size by methods of simplification or approximation in which very little of the original problem is lost. It is precisely such a cutting down or slicing up of the problem that we propose to treat in this paper, in the hope that it may help bring the theory and practice of statistical decision functions somewhat closer together.

The Modified Minimax Principle. The minimax principle, which looks only at the single value $\sup_\omega \rho(\omega, \varphi)$, suffers from the defect of being an over-simplification. Yet it suggests, by means of a simple modification, a natural way of approximating to the problem. This is to cut Ω up, to partition it into sets or "slices" Ω_s, s running over an index set S, and then look at $\sup_{\omega \in \Omega_s} \rho(\omega, \varphi) = \alpha(s, \varphi)$ for each s in S, so that corresponding to the partition $\Omega = \bigcup_{s \in S} \Omega_s$ we look only at the values $\alpha(s, \varphi)$ for $s \in S$ instead of at $\rho(\omega, \varphi)$ for all ω. The range of ρ_φ is thus replaced by the smaller range of α_φ, making comparisons of procedures that much easier. It is this slicing up of Ω (or its replacement by the smaller class S) and consequent simplification of the risk functions in the above way that we call the modified minimax principle. (By way of analogy, one might think of upper Darboux sums approximating Riemann integrals.) The reduced game can then be treated as any other game: one can play Bayes or minimax in it, or attempt to delineate its admissible strategies.

If the "slicing principle" used is such that the supremum of ρ_φ over each slice is not much different from the values within the slice, or has some other reasonable property, then very little is lost. The question of what are reasonable or natural slicing principles is clearly of primary importance here, and we shall present what we believe are several of them.

The theory of invariance provides us with the most powerful of these slicing principles and will play a central role in our considerations. The natural slicing of Ω into its orbits under the group leads us to what appears to be the best possible generalization of the Hunt-Stein theorem, and to its most natural setting: namely to the theorem that under certain regularity conditions which are often met in practice the invariant procedures form a complete class in the sense of the sliced up risk functions. The theory of composite hypotheses is also discussed in this light, and an example of theoretical interest is given illustrating these concepts, in which a difficult problem undergoes a striking simplification.

Finally, the use of previous experience as a slicing principle is discussed, and

related to a purely game-theoretic model which we have constructed for the modified minimax principle and which we have called a *mixed game*.

Though these slicing principles appear to be among the most important, the search for others continues. A method of approximation and simplification having been established, it remains to be seen whether these principles and available numerical methods can be combined to make an effective instrument in practice.

2. The use of invariance as a slicing principle. We begin with a brief description of the statistical and group-theoretic setup. Background material and further remarks on statistical games can be found in Blackwell and Girshick [1] and Wald [2]. For measure-theoretic and related notions, we refer to Halmos [3]. The formidability of notation, while unavoidable in this subject, is more apparent than real, and we shall try to clarify matters with a running commentary.

DEFINITION 1. By a statistical problem $(Z, \mathcal{B}, \Omega, P, A, \mathcal{A}, L)$ is meant

i) A set Z of all possible experimental outcomes, a Borel field $\mathcal{B}$ of subsets of Z, a set Ω of states of nature, and a function P defined on $\mathcal{B} \times \Omega$ such that for each $\omega \in \Omega$, P_ω is a probability measure on $(Z, \mathcal{B})$. $(Z, \mathcal{B}, \Omega, P)$ is called the sample space

ii) A set of actions A available to the statistician, together with a Borel field $\mathcal{A}$ of subsets of A

iii) A loss function L defined on $\Omega \times A$ such that $L(\omega, a)$ is the loss to the statistician when he takes action a and ω is the true state of nature. For each $\omega \in \Omega$, L_ω is assumed to be a non-negative $\mathcal{A}$-measurable function. (A mapping $f: M \to N$ of one measurable space $(M, \mathfrak{M})$ into another $(N, \mathfrak{N})$ is said to be $\mathfrak{N}$-$\mathfrak{M}$ measurable if the inverse image of every $\mathfrak{N}$-set is an $\mathfrak{M}$-set. If $(N, \mathfrak{N})$ is Euclidean, with the usual Borel sets, we call the mapping simply $\mathfrak{M}$-measurable. If E is a subset of M, we shall write $f''E$ for the image of E under f.)

DEFINITION 2. $(\mathcal{G}, \gamma_Z, \gamma_\Omega, \gamma_A)$ is said to be an admissible group on the statistical problem, or the problem is said to be invariant under the group $\mathcal{G}$ (for short) if

i) $\mathcal{G}$ is a group

ii) γ_Z is a homomorphism of $\mathcal{G}$ into the class of all $\mathcal{B}$-$\mathcal{B}$ measurable 1-1 transformations of Z onto itself, γ_Ω is a homomorphism of $\mathcal{G}$ into the class of all 1-1 transformations of Ω onto itself, γ_A is a homomorphism of $\mathcal{G}$ into the class of all $\mathcal{A}$-$\mathcal{A}$ measurable 1-1 transformations of A onto itself, where for each $g \in \mathcal{G}$ we shall write g_Z for $\gamma_Z(g)$, g_Ω for $\gamma_\Omega(g)$, and g_A for $\gamma_A(g)$

iii) for each $g \in \mathcal{G}$, $B \in \mathcal{B}$, $\omega \in \Omega$, we have $P(g_Z'' B \mid g_\Omega(\omega)) = P(B \mid \omega)$ and

iv) for each $g \in \mathcal{G}$, $\omega \in \Omega$, $a \in A$, we have $L(g_\Omega(\omega), g_A(a)) = L(\omega, a)$.

What this definition says is that each element g of group $\mathcal{G}$ is in effect three simultaneous permutations or relabelings, g_Z, g_Ω, and g_A, of the elements of Z, Ω, and A, respectively, together with their respective Borel fields, under which probabilities of sets and losses due to actions are invariant. The homomorphisms further imply that we have three groups $\mathcal{G}_Z$, $\mathcal{G}_\Omega$, and $\mathcal{G}_A$ of such permutations.

As a simple example illustrating these concepts, we may think of the following: Let Z be the real line, $\mathcal{B}$ the ordinary Borel sets, Ω the real line, P_ω the normal

distribution with mean ω and variance 1, $A = Z$, $\mathfrak{A} = \mathfrak{B}$, and $L(\omega, a) = \psi(\omega - a)$, i.e., the loss is some function depending only on the difference $\omega - a$. The problem is easily seen to be invariant under the full translation group $\mathcal{G}$ each element of which gives rise to three identical translations of the three real lines Z, Ω, A by the same real number. The reader may generalize this example at once to the n-dimensional case, with P_ω the multivariate normal distribution $N(\omega, I)$, vector $\omega = (\omega_1, \cdots, \omega_n)$ of means, covariance matrix I, and translations $z \to z + g = (z_1 + g, \cdots, z_n + g)$.

DEFINITION 3. A randomized decision procedure is a function φ on $\mathfrak{A} \times Z$ to the unit interval of reals $[0, 1]$ such that for each z, φ_z is a probability measure on $(A, \mathfrak{A})$, i.e., $\varphi(T \mid z)$ is the probability of taking an action in $T \varepsilon \mathfrak{A}$, given z, and such that for fixed T, φ_T is a $\mathfrak{B}$-measurable function of z. We write Φ for the class of all randomized decision functions.

DEFINITION 4. For any state of nature ω and decision procedure φ, the value of the risk function ρ is given by

$$\rho(\omega, \varphi) = \int\int L(\omega, a)\, d\varphi(a \mid z)\, dP_\omega(z).$$

DEFINITION 5. For each $g \varepsilon \mathcal{G}$, we make correspond to each procedure φ a new procedure $g_\Phi\varphi$ given by

$$(g_\Phi\varphi)(T \mid z) = \varphi(g_A''T \mid g_Z(z)).$$

It is an easy but very important consequence of our definitions that the risk function ρ is invariant, or as we say, transforms "correctly," under the group by means of the formula

$$\rho(\omega, g_\Phi\varphi) = \rho(g_\Omega(\omega), \varphi) \qquad \text{for every } g \varepsilon \mathcal{G}.$$

The simple verification of this is as follows:

$$\rho(\omega, g_\Phi\varphi) = \int\int L(\omega, a)\, d\varphi(g_A(a) \mid g_Z(z))\, dP(z \mid \omega).$$

Writing $L(\omega, a) = L(g_\Omega(\omega), g_A(a))$ and $dP(z \mid \omega) = dP(g_Z(z) \mid g_\Omega(\omega))$, then relabeling $g_Z z = z'$ and $g_A(a) = a'$, we get

$$= \int\int L(g_\Omega(\omega), a')\, d\varphi(a' \mid z')\, dP(z' \mid g_\Omega(\omega))$$

$$= \rho(g_\Omega(\omega), \varphi).$$

DEFINITION 6. A decision procedure φ is said to be invariant under the group if $g_\Phi\varphi = \varphi$ for every $g \varepsilon \mathcal{G}$, i.e., if

$$\varphi(g_A''T \mid g_Z(z)) = \varphi(T \mid z) \qquad \text{for all } g, z, T.$$

Note, then, that if φ is invariant, $\rho(\omega, \varphi) = \rho(g_\Omega(\omega), \varphi)$ for every $g \varepsilon \mathcal{G}$.

A brief word at this point on the effect in general of an arbitrary group of permutations on an arbitrary set will give some direction to these considerations

and will enable us to convey the essence and spirit of what is known as *the invariance principle.*

Let G be a group of permutations of the elements of a set X. The existence of a permutation sending one element of X into another is readily seen to establish an equivalence relation in X and gives rise to a useful partition of X into equivalence classes, known as "orbits," under the group. Formally, given $x \varepsilon X$, the orbit V_x to which it belongs is the set of elements y of X given by

$$V_x = \{y : y = gx \text{ for at least one } g \varepsilon G\}.$$

Now so far as the effect of G in its bearing on X is concerned, there is no distinguishing between elements of the same orbit. Taking an anthropomorphic view of the group, from the point of view of "the man in the group," all elements in an orbit look alike, the group in its dealings with the set displaying blindness to a mere matter of a difference in labels. To put it another way, when the man in the group looks at X, he sees only orbits.

The invariance principle. Returning to our statistical setup, the space Z of experimental outcomes breaks up into orbits under the influence of $\mathcal{G}$, more specifically under the permutation group $\mathcal{G}_Z$. Ω also breaks up into orbits under $\mathcal{G}_\Omega$. By the invariance principle is meant the adoption by the statistician of the viewpoint of the man in the group $\mathcal{G}$: he becomes himself the man in the group in that he too sees the problem in terms of orbits only. Specifically, this means that the only decision procedures he will employ are the invariant ones, that is to say, those procedures which, while free to vary at will from orbit to orbit of Z, must exhibit within each orbit complete consistency with respect to the groups $\mathcal{G}_Z$ and $\mathcal{G}_A$ as prescribed in Definition 6.

The main reason for restricting ourselves to invariant procedures has always been their undeniable plausibility. We are further encouraged by the fact that in many problems (under a finite group, say, and in other cases) an invariant decision procedure admissible within the class of all invariant procedures is known to be admissible among all procedures as well. Moreover, if we should be looking for minimax strategies, the Hunt-Stein theorem (to be described in the next section in terms of testing hypotheses) tells us that under certain weak assumptions a minimax invariant procedure is minimax among all procedures too. For us, however, there is now another and more compelling reason. It turns out that the use of invariance as a slicing principle, so that orbits rather than points become the basic elements under consideration, leads in our modified minimax sense to the best possible generalization of the Hunt-Stein result, namely, that the invariant procedures form a complete class in the sense of the sliced up risk functions. In the modified minimax sense, then, the use of invariance provides a striking simplification of the original problem, particularly when the orbits, hence the groups, are rather large.

It should be noted in this connection, by the remark in Definition 6 above, that risk functions ρ_φ for invariant procedures φ are constant over each orbit of Ω, as is to be expected, so that invariant procedures lose nothing under the modi-

fied minimax principle of taking suprema. Before presenting our main result in some detail, it will be convenient and appropriate to give a brief description of the invariance principle in the language of testing hypotheses, and to state the classical Hunt-Stein theorem in these terms.

3. The principle of invariance in testing hypotheses. Let $(Z, \mathfrak{B}, \Omega, P)$ with $\Omega = \Omega_0 \cup \Omega_1$ a disjoint union, be a sample space, and suppose we want to test $H_0 : \omega \in \Omega_0$ against $H_1 : \omega \in \Omega_1$.

Let there exist a σ-finite measure ν on $(Z, \mathfrak{B})$ such that there is a real-valued function p on $Z \times \Omega$ with

$$P(B \mid \omega) = \int_B p(z \mid \omega)\, d\nu(z) \qquad \text{for every } \omega \in \Omega \text{ and } B \in \mathfrak{B},$$

i.e., we are assuming the existence of a σ-finite measure ν which dominates all our probability distributions P_ω, and p_ω is the Radon-Nikodym derivative $dP_\omega/d\nu$.

DEFINITION 1. The group $\mathcal{G}$ is said to keep the testing problem invariant if $\omega \in \Omega_i$ implies $g_0(\omega) \in \Omega_i$, $i = 0, 1$, for every $g \in \mathcal{G}$, that is to say, if each permutation g_0 of the elements of Ω permutes the elements of Ω_0 and of Ω_1 separately. (Note that this implies for the testing problem that we are specializing the general statistical problem down to a two-action setup $A = \{0, 1\}$ with constant losses due to a wrong decision and $g_A = I$, the identity mapping, for every g.)

DEFINITION 2. By a randomized test is meant a $\mathfrak{B}$-measurable real-valued function φ on Z with values in the closed interval $[0, 1]$, where $\varphi(z)$ is the probability of rejecting H_0 when z is the observed sample point.

DEFINITION 3. The test φ is said to be invariant under $\mathcal{G}$ if for all g and z $\varphi(g_Z(z)) = \varphi(z)$ except at most on a set of ν-measure 0, i.e., if φ is essentially constant on the orbits of Z. If the exceptional set of ν-measure 0 depends on g, then φ is called almost invariant.

Let $\mathfrak{C}$ be a Borel field of subsets of $\mathcal{G}$. Two assumptions will always be made:

i) $(g, z) \to g_Z(z)$ is $\mathfrak{B} - \mathfrak{C} \times \mathfrak{B}$ measurable

ii) $(g_1, g_2) \to g_1 g_2$ is $\mathfrak{C} - \mathfrak{C}^2$ measurable.

DEFINITION 4. A measure μ on $(\mathcal{G}, \mathfrak{C})$ is said to be right invariant if $\mu(Cg) = \mu(C)$ for every $g \in \mathcal{G}$ and every $C \in \mathfrak{C}$. Left invariance is defined using $\mu(gC) = \mu(C)$.

As an example of an invariant measure we may take for $\mathcal{G}$ the additive group of reals, $\mathfrak{C}$ the ordinary Borel sets, and μ Lebesgue measure on the reals. No invariant probability measure, however, exists for this group, which gives rise to the following useful limiting notion.

DEFINITION 5. A sequence $\{\mu_n\}$ of probability measures on $(\mathcal{G}, \mathfrak{C})$ is said to be asymptotically right invariant if $\lim_{n\to\infty}(\mu_n(Cg) - \mu_n(C)) = 0$ for every $g \in \mathcal{G}$ and $C \in \mathfrak{C}$.

As an example we take the same group just given, and for each $C \in \mathfrak{C}$ and $n = 1, 2, \cdots$, define the probability measure μ_n on $(\mathcal{G}, \mathfrak{C})$ by

$$\mu_n(C) = (1/2n)\mu(C \cap [-n, n]),$$

i.e., we approximate Lebesgue measure μ by the conditional probabilities μ_n given that g belongs to the closed interval $[-n, n]$. Clearly the μ_n are all probability measures and $1/2n$ is the normalizing factor. To prove asymptotic invariance, we have, since clearly $Cg \cap [-n, n] = C \cap [-n - g, n - g]$,

$$|\mu_n(Cg) - \mu_n(C)| = \left|\frac{1}{2n}\mu(C \cap [-n - g, n - g]) - \frac{1}{2n}\mu(C \cap [-n, n])\right|$$

$$\leqq \frac{|g|}{2n} \to 0 \qquad \text{as } n \to \infty,$$

showing in fact that the approach to zero for each g is here even uniform in C, that is,

$$\lim_{n\to\infty} \sup_{C} |\mu_n(Cg) - \mu_n(C)| = 0 \qquad \text{for every } g \in \mathcal{G}.$$

(This example generalizes at once to the additive group $\mathcal{G}$ (real p-tuples) of a p-dimensional real linear space. One takes conditional probabilities of Lebesgue measure given the cubes $[(-n, \cdots, -n); (n, \cdots, n)]$, normalizing by $1/(2n)^p$, and, writing $g = (g_1, \cdots, g_p)$, showing that for n sufficiently large, in fact for $2n > \max_{i=1,\cdots,p} |g_i|$,

$$|\mu_n(Cg) - \mu_n(C)| \leqq 1 - \prod_{i=1}^{p}\left(1 - \frac{|g_i|}{2n}\right) \to 0 \qquad \text{as } n \to \infty.$$

The Hunt-Stein theorem [4], in the language of testing problems, then reads:

THEOREM. *Let there exist, for the testing problem as just defined, an asymptotically (right) invariant sequence of probability measures $\{\mu_n\}$ on $(\mathcal{G}, \mathcal{C})$. Then there exists an invariant test φ_0 among all tests φ for which $\int\varphi(z)p(z \mid \omega)\,d\nu(z) \leqq \alpha$ for all $\omega \in \Omega_0$, and which maximizes $\inf_{\omega\in\Omega_1}\int\varphi(z)p(z \mid \omega)\,d\nu(z)$; i.e., there exists an invariant minimax test φ_0.* (The probability with which a test φ rejects H_0 when ω is the true state of nature is given by $\beta_\varphi(\omega) = \int\varphi(z)p(z \mid \omega)\,d\nu(z)$. β_φ is called the *power function* of the test and is the only property of the test that interests us, bearing an obvious relation to the risk function. The original version of the Hunt-Stein theorem was in terms of most stringent tests. The connection with this version is easily seen if we subtract one minus the size α envelope power function $\beta_\alpha^*(\omega)$ from the loss function for $\omega \in \Omega_1$. The maximum risk for $\omega \in \Omega_1$ is then the stringency.)

It will be very helpful for our later understanding to present a brief sketch of the proof for this simple case, as it contains the essential ideas without the technical difficulties of the generalization. It depends rather heavily on the fact that the space of all randomized tests φ is compact in the weak -* topology, or more simply upon one version of this which is the following well-known lemma due to F. Riesz (see Banach [5] page 130).

LEMMA OF F. RIESZ. *If $\{\varphi_n\}$ is a sequence of $\mathfrak{B}$-measurable functions on Z to the closed interval $[0, 1]$—a sequence of randomized tests in our terminology—then there*

exists a subsequence $\{\varphi_{n_i}\}$ *and a test* φ' *such that for every function* f *integrable with respect to* ν

$$\lim_{i\to\infty}\int \varphi_{n_i}(z)f(z)\,d\nu(z) = \int \varphi'(z)f(z)\,d\nu(z).$$

PROOF OF THE HUNT-STEIN THEOREM: The first step of the proof consists in showing, by a simple application of Riesz's lemma, that a minimax solution φ_0 exists. (It is in fact the φ' of the lemma associated with a sequence of tests approaching the minimax condition.) For every $g \in \mathcal{G}$, then, we have that the test $g_*\varphi_0$ given by $(g_*\varphi_0)(z) = \varphi_0(g_Z(z))$ is also minimax, since the invariance of the risk function simply means here that the values of the power function of $g_*\varphi_0$ are just those of the power function of φ_0 permuted separately over Ω_0 and over Ω_1. The minimax tests $g_*\varphi_0$ are then averaged out by the probability measures μ_n on $(\mathcal{G}, \mathfrak{C})$, giving rise to the sequence of tests

$$\varphi_n(z) = \int_{\mathcal{G}} \varphi_0(g_Z(z))\,d\mu_n(g).$$

These tests φ_n are again minimax, as a simple examination of their power functions shows (interchanging orders of integration by Fubini's theorem, we get, for each ω, permuted values of power functions averaged out over $\mathcal{G}$ by μ_n). Now take $\{\varphi_{n_i}\}$ and φ_0' in the sense of Riesz's lemma. Then just as for φ_0 above, φ_0' is also a minimax test. The one remaining difficulty is to show that φ_0' is an invariant test, more precisely an almost invariant test, and this is accomplished in two steps by a straightforward integration argument:

1) To show, for each $g \in \mathcal{G}$, that $\varphi_0'(g_Z(z)) = \varphi_0'(z)$ for almost all $z[\nu]$, it suffices to show that $\int\varphi_0'(g_Z(z))f(z)\,d\nu(z) = \int\varphi_0'(z)f(z)\,d\nu(z)$ for every integrable function f,

2) and this readily reduces to showing that a consequence of asymptotic invariance of a sequence $\{\mu_n\}$ of probability measures is that

$$\lim_{n\to\infty}\left[\int_{g'\in\mathcal{G}} \psi(g')\,d\mu_n(g'\cdot g^{-1}) - \int_{g'\in\mathcal{G}} \psi(g')\,d\mu_n(g')\right] = 0$$

for every bounded measurable function ψ of g'.

Having done this, φ_0' is almost invariant. Finally, a standard argument allows us to replace φ_0' by an invariant test φ_0'' such that $\varphi_0'' = \varphi_0'$ almost everywhere with respect to ν, provided only that another regularity condition is imposed on the group $\mathcal{G}$.

4. The modified minimax principle, completeness, and the generalized Hunt-Stein theorem. We return in this section to the general setup of a statistical decision problem invariant under a group of transformations, as described in Section 2, and assume in addition that all our probability distributions P_ω are dominated by σ-finite measure ν on $(Z, \mathfrak{B})$. Without loss of generality, see Halmos and Savage [6], ν may be assumed equivalent to the family of all P_ω, that is, if $P_\omega(B) = 0$ for all ω, then $\nu(B) = 0$.

From Definitions 3 and 4 of Section 2, it is clear that a decision procedure φ can be changed on a set of ν-measure 0 without changing its risk function, so that two procedures differing on such a set may henceforth always be taken as equivalent. If in Definition 6 we have $g_{\#}\varphi = \varphi$ for each $g \in \mathcal{G}$, where the exceptional set of ν-measure 0 depends on g, we shall call φ almost invariant. We denote the class of almost invariant procedures by Φ^* and the class of invariant procedures by Φ^{**}. Further, we write $\Omega = \cup_{s \in S}\Omega_s$ where the Ω_s, s running over an index set S, are the orbits of Ω under the group $\mathcal{G}_\Omega$.

In strict accordance with our modified minimax principle as described in Sections 1 and 2, we have the following definition:

DEFINITION 1. A procedure φ is said to be at least as good in the modified minimax sense as a procedure ψ if

$$\sup_{\omega \in \Omega_s} \rho(\omega, \varphi) \leqq \sup_{\omega \in \Omega_s} \rho(\omega, \psi) \qquad \text{for every } s \in S.$$

The related notions of admissibility, complete class, etc., in the modified minimax sense are all similarly defined, in the obvious way. Henceforth in this section, such phrases as "at least as good as" are always to be understood in the modified minimax sense.

In generalizing upon the Hunt-Stein theorem, we run into the complicated notions of weak -* compactness, convergence and cluster point for a set of measures on a topological space. As these concepts are by no means obvious, it will be necessary to spell them out in some detail. Readers who wish to avoid topological difficulties may pass lightly over this part of the proof, pausing only long enough to note that in substance it is designed to show how a cluster point of a sequence of procedures may be used in constructing a new decision procedure capable of playing the central role analogous to the one in Riesz's lemma.

With the general plan of the proof of the Hunt-Stein theorem as outlined in Section 3 firmly in mind, we now proceed to our generalization. It will be useful to make here the obvious generalization of Definition 5, Section 3, to an asymptotically right invariant net $\{\mu_\alpha\}$ of probability measures on $(\mathcal{G}, \mathcal{C})$. Readers who are unfamiliar with the notion of nets may continue to think of sequences. For a discussion of nets we refer to Kelley [7]. The modified minimax principle, it should be remembered, requires us to pay strict attention to the fact that the various permutations mentioned there never take us out of an orbit, leading to the sharper result of completeness.

THE GENERALIZED HUNT-STEIN THEOREM. *Let* $(Z, \mathcal{B}, \Omega, P, A, \mathcal{A}, L)$ *be a statistical problem invariant under a group* $\mathcal{G}$ *as previously described. If the following regularity conditions are satisfied,*

i) *there exists a* σ*-finite measure* ν *equivalent to the set of all* P_ω,

ii) *there is a topology on the action set* A *such that* A *is a separable metric space and such that* $\mathcal{A}$ *is the Borel field generated by the compact subsets of* A, *and the loss function* L *is such that for each* $\omega \in \Omega$, $L(\omega, a)$ *is non-negative and continuous in* a *and is such that for every real number* τ, *the set* $\{a : L(\omega, a) \leqq \tau\}$ *is compact,*

iii) $\mathcal{C}$ *is a Borel field of subsets of* $\mathcal{G}$ *and there is an asymptotically right invariant net* $\{\mu_\alpha\}$ *of probability measures on* $(\mathcal{G}, \mathcal{C})$,
then the class Φ^* *of almost invariant decision procedures is essentially complete in the modified minimax sense. If in addition*

iv) $\mathcal{G}$ *is a locally compact* σ*-compact topological group with* $\mathcal{C}$ *generated by the compact subsets of* $\mathcal{G}$,
then the class Φ^{**} *itself is essentially complete in this sense.*

PROOF. Let $\varphi \in \Phi$ be given. We are required to find a $\varphi^* \in \Phi^*$ at least as good as φ. We may assume without loss of generality that the risks of φ are bounded on each orbit Ω_α by finite numbers m_α, for otherwise we may clearly ignore such orbits when comparing a possible φ^* and have only to carry out the following argument for the remaining orbits.

A) We observe first of all that for each $g \in \mathcal{G}$, the procedure $g_\natural\varphi$ is equivalent to φ in the modified minimax sense. This follows at once from the invariance of the risk function, expressed by the formula $\rho(\omega, g_\natural\varphi) = \rho(g_\Omega(\omega), \varphi)$, which tells us that the values of one risk function are simply those of the other permuted within each orbit of Ω separately.

We define for each α,

$$\varphi_\alpha(T \mid z) = \int_{\mathcal{G}} \varphi(g''_A T \mid g_Z(z))\, d\mu_\alpha(g),$$

a net of randomized procedures got by averaging the procedures $g_\natural\varphi$ with respect to the measures μ_α on $(\mathcal{G}, \mathcal{C})$, and shall show by a simple calculation of their risk functions that the φ_α are all at least as good as φ.

$$\rho(\omega, \varphi_\alpha) = \int\int L(\omega, a)\, d\varphi_\alpha(a \mid z)\, dP_\omega(z).$$

The Lebesgue integral transformation comes to our aid here, enabling us to compute such an expression by writing it as

$$= \int\int_0^\infty \varphi_\alpha(\{a : L(\omega, a) > h\} \mid z)\, dh\, dP_\omega(z)$$

$$= \int\int\int \varphi(g''_A\{a : L(\omega, a) > h\} \mid g_Z(z))\, d\mu_\alpha(g)\, dh\, dP_\omega(z).$$

Now interchanging orders of integration (the integrand is positive) we integrate out with respect to h and the Lebesgue integral transformation in reverse gives us

$$= \int\int\int L(\omega, a)\, d\varphi(g_A(a) \mid g_Z(z))\, dP_\omega(z)\, d\mu_\alpha(g)$$

$$= \int \rho(\omega, g_\natural\varphi)\, d\mu_\alpha(g)$$

$$= \int \rho(g_\Omega(\omega), \varphi)\, d\mu_\alpha(g)$$

$$\leqq \sup_{\omega' \in \Omega_\alpha} \rho(\omega', \varphi),$$

where Ω_s is the particular orbit to which ω belongs; hence,

$$\sup_{\omega \in \Omega_s} \rho(\omega, \varphi_\alpha) \leqq \sup_{\omega \in \Omega_s} \rho(\omega, \varphi) \qquad \text{for all } s \in S,$$

so that the φ_α are all at least as good as φ.

B) Let $\mathfrak{K}$ be the collection of compact subsets K of A. For fixed K and any $\varphi \in \Phi$, φ_K is a $\mathfrak{B}$-measurable function on Z to the closed interval $[0, 1]$. By Φ_K we mean the collection of all these functions φ_K with $\varphi \in \Phi$, i.e., the set of all randomized tests. Now the Banach space $L^\infty(\nu)$ of all bounded $\mathfrak{B}$-measurable functions is the adjoint of the Banach space $L^1(\nu)$ of all extended real-valued functions integrable with respect to ν, and Alaoglu's theorem tells us that the solid unit sphere of the adjoint of a Banach space is compact in the weak -* topology. Hence Φ_K, as the intersection of a closed set with the solid unit sphere, is also compact in the weak -* topology on $L^\infty(\nu)$. By Tychonoff's compactness theorem, the cartesian product of the compact sets Φ_K for all $K \in \mathfrak{K}$ is a compact set in the product topology of these Banach spaces $L^\infty(\nu)$. Therefore our net of functions φ_α defined above has a cluster point $\bar{\varphi}$ in the product space. A consequence of this, called the cluster point condition, is that for any finite number of sets K, any finite number of functions f integrable with respect to ν and any number $\epsilon > 0$, there is an arbitrarily large α such that

$$\left| \int \varphi_\alpha(K \mid z) f(z)\, d\nu(z) - \int \bar{\varphi}(K \mid z) f(z)\, d\nu(z) \right| < \epsilon$$

Further, since $\bar{\varphi}$ is a cluster point of the net of functions φ_α, there is a subnet φ_{α_β} converging to it. To simplify the argument, we shall henceforth confine ourselves to this subnet and drop the second subscript β, so that $\bar{\varphi}$ is now a limit point of the $\{\varphi_\alpha\}$, i.e., for any compact set K and any integrable f we have

$$\int \varphi_\alpha(K \mid z) f(z)\, d\nu(z) \to \int \bar{\varphi}(K \mid z) f(z)\, d\nu(z).$$

Note that nothing is said about $\bar{\varphi}$ being a procedure, for it most certainly need not be. However, we have the following. By definition, every procedure φ has the properties

1) If $K_1 \subset K_2$, then $\varphi_{K_1} \leqq \varphi_{K_2}$;
2) If $K_1 \cap K_2 = \phi$, then $\varphi_{K_1 \cup K_2} = \varphi_{K_1} + \varphi_{K_2}$;
3) $\varphi_K \leqq 1$;

and it is easily seen that the limit point $\bar{\varphi}$ must have the same three properties almost everywhere with respect to ν (in fact has the third property everywhere). For example, we show property 2).

Let f be a function everywhere positive and integrable with respect to ν (such an f clearly exists since ν is σ-finite). Define a function g by

$$g(z) = f(z)[\bar{\varphi}(K_1 \cup K_2 \mid z) - \bar{\varphi}(K_1 \mid z) - \bar{\varphi}(K_2 \mid z)].$$

g is still integrable, as the product of f by a bounded $\mathfrak{B}$-measurable function. Applying the limit point condition to the function g and the three sets $K_1 \cup K_2$,

K_1, K_2 we have

$$\int g(z)[\bar{\varphi}(K_1 \cup K_2 \mid z) - \bar{\varphi}(K_1 \mid z) - \bar{\varphi}(K_2 \mid z)]\, d\nu(z) - \int g(z)[\varphi_\alpha(K_1 \cup K_2 \mid z) - \varphi_\alpha(K_1 \mid z) - \varphi_\alpha(K_2 \mid z)]\, d\nu(z) \to 0.$$

But the second square bracket is zero, hence rewriting g

$$\int f(z)[\bar{\varphi}(K_1 \cup K_2 \mid z) - \bar{\varphi}(K_1 \mid z) - \bar{\varphi}(K_2 \mid z)]^2\, d\nu(z) = 0.$$

But f was chosen everywhere positive, hence

$$\bar{\varphi}_{K_1 \cup K_2} = \bar{\varphi}_{K_1} + \bar{\varphi}_{K_2}[\nu].$$

We now show how $\bar{\varphi}$ may be used to construct a suitable procedure φ^*.

Let $\mathfrak{R}$ be a countable subset of $\mathfrak{K}$ such that $\mathfrak{R}$ is closed with respect to finite unions and finite intersections and such that every open subset U of A is a countable union of interiors of elements R of $\mathfrak{R}$ which are themselves subsets of U (such an $\mathfrak{R}$ is known to exist by virtue of our hypothesis (ii) on A). Consider the function $\bar{\varphi}$ restricted to the set $\mathfrak{R}$. Thus restricted, the limit point $\bar{\varphi}$ has the three mentioned properties of a φ with only countably many exceptional sets of ν-measure 0, which we may hereby disregard and assume the properties hold everywhere for $\bar{\varphi}$.

We define an outer measure φ^* on the open sets U in A by the formula, for each z,

$$\varphi^*(U \mid z) = \sup_{R \subset U} \bar{\varphi}(R \mid z).$$

φ^*_U is clearly a $\mathfrak{B}$-measurable function of z, and for each z we do have an outer measure by the subadditivity of the supremum. Hence, we can extend φ^* to be a Caratheodory outer measure by defining for every subset W of A,

$$\varphi^*(W \mid z) = \inf_{U \supset W} \varphi^*(U \mid z).$$

φ^*_W is clearly $\mathfrak{B}$-measurable, and for each z all the Borel sets $T \in \mathfrak{A}$ are measurable. Restricting φ^* to the set $\mathfrak{A}$, it is thus a measure on $(A, \mathfrak{A})$ for each z. If we can show, therefore, that $\varphi^*(A \mid z) = 1$ $[\nu]$, φ^* will then be a randomized decision procedure. To do this, we first show that for every compact set K, $\varphi^*_K \geqq \bar{\varphi}_K[\nu]$:

Let $U \supset K$ be any open set. Then we have seen U may be written as $U = \bigcup_{R_i \subset U} (\text{interior of } R_i)$ which is then an open covering for the compact set K, so that for some n, $K \subset \bigcup_{i=1}^n R_i = R \in \mathfrak{R}$. Hence $\bar{\varphi}_K \leqq \bar{\varphi}_R$ $[\nu]$. But $R \subset U$, so that $\bar{\varphi}_K \leqq \varphi^*_U[\nu]$. Now write $K = \bigcap_j U_j\downarrow$ the intersection of a descending sequence of open sets, which can be done since our space is metric. Since φ^* is a finite measure on the Borel sets, we have $\varphi^*_K = \lim \varphi^*_{U_j}$ everywhere. Combining this with the last inequality gives $\varphi^*_K \geqq \bar{\varphi}_K[\nu]$ since there are only countably many j involved.

Now consider any $\epsilon > 0$ and any $\omega \in \Omega$. We have for every α, $\rho(\omega, \varphi_\alpha) \leqq m_\epsilon$,

where Ω_ω is the particular orbit to which ω belongs. A simple consequence of this inequality is that for the compact set $K = \{a:L(\omega, a) \leqq m_\omega/\epsilon\}$ and for every α there holds

$$\int \varphi_\alpha(K \mid z)\, dP_\omega(z) > 1 - \epsilon.$$

Hence by the limit point condition, since $dP_\omega(z) = p(z \mid \omega)\, d\nu(z)$, we have

$$\int \bar{\varphi}(K \mid z)\, dP_\omega(z) \geqq 1 - \epsilon,$$

so that by the inequality just proved,

$$\int \varphi^*(K \mid z)\, dP_\omega(z) \geqq 1 - \epsilon,$$

therefore,

$$\int \varphi^*(A \mid z)\, dP_\omega(z) \geqq 1 - \epsilon.$$

But ϵ is arbitrary, and as A is open, we know $\varphi^*(A \mid z) \leqq 1$ everywhere; hence, $\varphi^*(A \mid z) = 1$ almost everywhere with respect to P_ω, for each $\omega \in \Omega$. From the equivalence of ν to the family of probability measures P_ω, there clearly follows $\varphi^*(A \mid z) = 1$ [ν].

We might mention in passing that this newly constructed procedure φ^* is a limit point for the net of procedures φ_α when the appropriate (weak -*) topology is put on the space Φ of randomized decision procedures, so that we have shown that the set of all procedures having risk $\leqq m_\omega$ for each ω, where the m_ω are preassigned finite numbers, is compact in the weak -* topology. [The appropriate weak -* topology for procedures is characterized as follows. A procedure φ is a limit point of a net $\{\varphi_\alpha\}$ of procedures if for every compact set K and every non-negative integrable function f, there holds the inequality $\lim\sup_\alpha \int \varphi_\alpha(K \mid z) f(z)\, d\nu(z) \leqq \int \varphi(K \mid z) f(z)\, d\nu(z)$. Since $\lim_\alpha \int \varphi_{\alpha K} f\, d\nu = \int \bar{\varphi}_K f\, d\nu$ and $\bar{\varphi}_K \leqq \varphi^*_K[\nu]$ and $f \geqq 0$, the inequality $\lim_\alpha \int \varphi_{\alpha K} f\, d\nu \leqq \int \varphi^*_K f\, d\nu$ follows.]

C) We now show that this new procedure φ^* just constructed satisfies the conclusion of the theorem.

We consider its risk function, making the Lebesgue integral transformation and interchanging orders of integration when necessary.

$$\begin{aligned}
\rho(\omega, \varphi^*) &= \int\int \varphi^*(\{a:L(\omega, a) > h\} \mid z) p(z \mid \omega)\, d\nu(z)\, dh \\
&= \int\int [1 - \varphi^*(\{a:L(\omega, a) \leqq h\} \mid z)] p(z \mid \omega)\, d\nu(z)\, dh \\
&= \lim_{H \to \infty} \left(H - \int_0^H \int \varphi^*(\{a:L(\omega, a) \leqq h\} \mid z) p(z \mid \omega)\, d\nu(z)\, dh \right)
\end{aligned}$$

$$\leqq \lim_{H\to\infty}\left(H - \int_0^H \lim_{\alpha}\int \varphi_\alpha(\{a: L(\omega, a) \leqq h\} \mid z)p(z \mid \omega)\, d\nu(z)\, dh\right)$$

$$= \lim_{H\to\infty}\lim_{\alpha}\left(H - \int_0^H \int \varphi_\alpha(\{a: L(\omega, a) \leqq h\} \mid z)p(z \mid \omega)\, d\nu(z)\, dh\right)$$

$$= \lim_{H\to\infty}\lim_{\alpha}\int_0^H \int [1 - \varphi_\alpha(\{a: L(\omega, a) \leqq h\} \mid z)]p(z \mid \omega)\, d\nu(z)\, dh$$

$$\leqq \lim_{\alpha}\int\int \varphi_\alpha(\{a: L(\omega, a) > h\} \mid z)p(z \mid \omega)\, d\nu(z)\, dh$$

$$= \lim_{\alpha} \rho(\omega, \varphi_\alpha)$$

$$\leqq \sup_{\omega' \in \Omega_\omega} \rho(\omega', \varphi),$$

where Ω_ω is the particular orbit to which ω belongs, hence φ^* is at least as good as φ in the modified minimax sense.

It remains to show that φ^* is an almost invariant procedure. To do this, we shall require the following lemma.

LEMMA. *The asymptotic right invariance of* $\{\mu_\alpha\}$ *implies for fixed* $g \in \mathcal{G}$ *that*

$$\lim_{\alpha}\int_{g' \in \mathcal{G}} \psi(g')[d\mu_\alpha(g' \cdot g^{-1}) - d\mu_\alpha(g')] = 0$$

for every bounded measurable function ψ *of* g'.

PROOF. This follows directly from the definition of integral. Without loss of generality, we assume $|\psi| \leqq 1$. Partitioning the interval $[-1, 1]$ into equal subintervals of length $1/M$, we have at once for each positive integer M the inequality.

$$\left|\int \psi(g')[d\mu_\alpha(g' \cdot g^{-1}) - d\mu_\alpha(g')]\right.$$
$$\left. - \sum_{k=-M}^{M} \frac{k}{M}\left[\mu_\alpha\left\{\left(g' : \frac{k}{M} \leqq \psi(g') < \frac{k+1}{M}\right) g^{-1}\right\}\right.\right.$$
$$\left.\left. - \mu_\alpha\left\{g' : \frac{k}{M} \leqq \psi(g') < \frac{k+1}{M}\right\}\right]\right| \leqq \frac{2}{M}.$$

For fixed M, the second square bracket approaches zero as α approaches infinity, hence $\lim\sup_\alpha |\int \psi(g')[d\mu_\alpha(g' \cdot g^{-1}) - d\mu_\alpha(g')]| \leqq 2/M$. As this is true for every M, the lemma follows.

Let, now, $g \in \mathcal{G}$ and a compact set K be given. We shall first show that $\bar{\varphi}(g''_A K \mid g_Z(z)) = \bar{\varphi}(K \mid z)[\nu]$. To do this, it suffices to show that

$$\int [\bar{\varphi}(g''_A K \mid g_Z(z)) - \bar{\varphi}(K \mid z)]f(z)\, d\nu(z) = 0$$

for every integrable function f. Since $\bar{\varphi}$ is a limit point of the φ_α in the weak -*

sense, the formula for φ_α in terms of φ makes this last integral

$$= \lim_\alpha \int\int \varphi(g_A'''K \mid g_x'(z))f(z)\, d\nu(z)[d\mu_\alpha(g' \cdot g^{-1}) - d\mu_\alpha(g')].$$

But $\psi(g') = \int \varphi(g_A'''K \mid g_x'(z))f(z)\, d\nu(z)$ is a measurable function of g' and is bounded by $\int |f|\, d\nu$, so our integral is indeed zero by our lemma.

We now show that $\varphi^*(g_A'' K \mid g_x(z)) = \varphi^*(K \mid z)[\nu]$. Let U, an open set, and L, a compact set, be given such that $K \subset U \subset L$. φ_U^* is by definition the supremum of the $\bar{\varphi}_R$, all of which are less than or equal to $\bar{\varphi}_L$, so that $\varphi_U^* \leqq \bar{\varphi}_L[\nu]$, thus permitting us to write $\varphi_K^* \leqq \varphi_U^* \leqq \bar{\varphi}_L \leqq \varphi_L^*[\nu]$. Applying the group element g to these inequalities, we may extract the following:

$$\varphi^*(g_A'' K \mid g_x(z)) \leqq \bar{\varphi}(g_A'' L \mid g_x(z)) = \bar{\varphi}(L \mid z) \leqq \varphi^*(L \mid z)\ [\nu].$$

We may also extract

$$\varphi^*(K \mid z) \leqq \bar{\varphi}(L \mid z) = \bar{\varphi}(g_A'' L \mid g_x(z)) \leqq \varphi^*(g_A'' L \mid g_x(z))\ [\nu].$$

Taking a sequence $\{L_n\}$ of such compact sets with $K = \bigcap L_n$ and passing to the limit we get $\varphi^*(g_A'' K \mid g_x(z)) \leqq \varphi^*(K \mid z) \leqq \varphi^*(g_A'' K \mid g_x(z))\ [\nu]$, thus establishing the equality $[\nu]$.

From the nature of our topology on A and the way $\mathfrak{A}$ is generated by the compact sets K, this last equality extends by a standard transfinite induction argument to every set $T \in \mathfrak{A}$, thus establishing the almost invariance of φ^*. The argument goes as follows.

Note first that $\mathfrak{A}$ is equal to the smallest monotone class $\mathfrak{M}(\mathfrak{K})$ containing the compact sets: A is σ-compact (the union of a sequence of compact sets). The σ-ring of subsets of A generated by $\mathfrak{K}$, containing the whole space A, is therefore a Borel field, thus equal to the Borel field $\mathfrak{A}$ generated by $\mathfrak{K}$. Obviously $\mathfrak{M}(\mathfrak{K}) \subset \mathfrak{A}$. The proper difference $K - L$ of two compact sets belongs to $\mathfrak{M}(\mathfrak{K})$ since, writing $L = \bigcap_n U_n$ the intersection of open sets, we have $K - L = \bigcup_n (K - U_n)$ which is σ-compact. Therefore the class of all finite unions of these proper differences belongs to $\mathfrak{M}(\mathfrak{K})$. Thus (see Halmos [3] Theorem F, p. 223) $\mathfrak{M}(\mathfrak{K})$ contains a ring containing $\mathfrak{K}$, so that (Halmos again, Theorem B, p. 27) $\mathfrak{M}(\mathfrak{K}) \supset \mathfrak{A}$, hence $\mathfrak{M}(\mathfrak{K}) = \mathfrak{A}$.

Write $\mathfrak{B}_0 = \mathfrak{K} = \{\text{the compact sets}\}$ and for every ordinal number α let $\mathfrak{B}_\alpha = \{\text{all monotone limits of sets } S_n, \text{ where } S_n \in \mathfrak{B}_{\beta_n} \text{ and } \beta_n < \alpha\}$. For every α, $\mathfrak{B}_\alpha$ is well-defined. Then $\mathfrak{A} = \mathfrak{M}(\mathfrak{K})$ means that $\mathfrak{A} = \bigcup_{\alpha < \omega_1} \mathfrak{B}_\alpha$ where ω_1 is the first uncountable ordinal. Having thus characterized $\mathfrak{A}$, the proof proceeds by transfinite induction on α: Suppose our equality $[\nu]$ holds for all $\mathfrak{B}_\beta$ with $\beta < \alpha$. Consider any set $T \in \mathfrak{B}_\alpha$. Then $T = \lim T_n$ where $T_n \in \mathfrak{B}_{\beta_n}$ and $\beta_n < \alpha$. The equality $[\nu]$ holds for these T_n by the induction hypothesis, therefore it holds for the limit T since there are only a countable number of sets of ν-measure 0 involved. The equality $[\nu]$ holding for all $T \in \mathfrak{B}_\alpha$, for all α, it holds for every $T \in \mathfrak{A}$, and φ^* is almost invariant.

D) With our assumptions on $(\mathfrak{G}, \mathfrak{C})$ we now proceed to replace this almost

invariant procedure φ^* by an invariant procedure φ^{**} such that $\varphi^{**} = \varphi^*$ [ν], thereby completing the proof of the theorem.

Let us return to the class $\mathfrak{R}$ of part (B). Since $\mathcal{G}$ is a locally compact topological group, there is a right invariant Haar measure λ on $(\mathcal{G}, \mathcal{C})$. We define a function φ^{**} as follows: For each $R \in \mathfrak{R}$

i) let $\varphi^{**}(R \mid z) = \varphi^*(R \mid z)$, for all z such that

$$\varphi^*(g''_\Delta R \mid g_x(z)) = \varphi^*(R \mid z) \quad [\lambda];$$

ii) for all such z in (i), and each g, let

$$\varphi^{**}(g''_\Delta R \mid g_x(z)) = \varphi^*(R \mid z);$$

iii) let $\varphi^{**}(R \mid z) = 0$ elsewhere.

To show that (i) and (ii) are not contradictory, suppose there existed R_1, z_1, g_1 with $R_2 = g''_{1_\Delta} R_1$, $z_2 = g_{1_x}(z_1)$, and

$$\varphi^*(g''_\Delta R_1 \mid g_x(z_1)) = \varphi^*(R_1 \mid z_1) \qquad [\lambda]$$

$$\varphi^*(g''_\Delta R_2 \mid g_x(z_2)) = \varphi^*(R_2 \mid z_2) \qquad [\lambda],$$

but with $\varphi^*(R_1 \mid z_1) \neq \varphi^*(R_2 \mid z_2)$. Then it would follow that

$$\begin{aligned} 0 &= \lambda\{g : \varphi^*(g''_\Delta R_2 \mid g_x(z_2)) \neq \varphi^*(R_2 \mid z_2)\} \\ &\geqq \lambda\{g : \varphi^*(g''_\Delta R_2 \mid g_x(z_2)) = \varphi^*(R_1 \mid z_1)\} \\ &= \lambda\{g : \varphi^*(g''_\Delta g''_{1_\Delta} R_1 \mid g_x(g_{1_x}(z_1))) = \varphi^*(R_1 \mid z_1)\}, \end{aligned}$$

which by the right invariance of λ

$$\begin{aligned} &= \lambda\{g : \varphi^*(g''_\Delta R_1 \mid g_x(z_1)) = \varphi^*(R_1 \mid z_1)\} \\ &= \lambda(\mathcal{G}). \end{aligned}$$

But $0 \geqq \lambda(\mathcal{G})$ contradicts the fact that λ is a Haar measure, hence φ^{**} is not contradictorily defined.

We now show, for each $R \in \mathfrak{R}$, that

$$\varphi^{**}(R \mid z) = \varphi^*(R \mid z) \quad [\nu]$$

Let

$$\begin{aligned} V &= \{(z, g) : \varphi^*(g''_\Delta R \mid g_x(z)) \neq \varphi^*(R \mid z)\} \\ V_z &= \{g : \varphi^*(g''_\Delta R \mid g_x(z)) \neq \varphi^*(R \mid z)\} \\ V_g &= \{z : \varphi^*(g''_\Delta R \mid g_x(z)) \neq \varphi^*(R \mid z)\}. \end{aligned}$$

The almost invariance of φ^* implies that $\nu(V_g) = 0$ for each g. Moreover, λ is σ-finite since $\mathcal{G}$ was assumed σ-compact, hence by Fubini's theorem (Halmos [3] Theorem A, p. 147) $(\nu \times \lambda)(V) = 0$. By the same reference this implies $\lambda(V_z) = 0$ [ν], hence $\varphi^{**}(R \mid z) = \varphi^*(R \mid z)$[$\nu$].

Since $\mathfrak{R}$ is countable, we may disregard sets of ν-measure 0, so that for all ...

z and g we have

$$\varphi^{**}(g''_A R \mid g_z(z)) = \varphi^*(R \mid z) = \varphi^{**}(R \mid z).$$

Extending this equality in the obvious way to the open sets U, then to the compact sets K, a transfinite induction on α exactly as in part (C) above extends it to all the Borel sets $T \varepsilon \mathfrak{A}$, thus completing the proof of the generalized Hunt-Stein theorem.

Remarks. It may prove useful in certain applications to point out an easy extension of the above result. Clearly, everything we have said goes through if Φ is any invariant, convex and closed set of randomized decision procedures rather than the class of all of them. That is, under our regularity conditions, if Φ is closed in the weak -* topology, and if for every probability measure λ on $(\mathcal{G}, \mathcal{C})$ and every $\varphi \varepsilon \Phi$, the procedure φ_λ given by $\varphi_\lambda(T \mid z) = \int \varphi(g''_A T \mid g_z(z))\, d\lambda(g)$ belongs to Φ, then for each $\varphi \varepsilon \Phi$ there is an almost invariant $\varphi^* \varepsilon \Phi$ which is at least as good as φ in the modified minimax sense, and so on. For example, the classical Hunt-Stein theorem follows from this by taking Φ to be the set of all procedures φ satisfying the inequality $\sup_{\omega \varepsilon \Omega_0} \rho(\omega, \varphi) \leqq \alpha$.

The condition in hypothesis (ii) that $L(\omega, a)$ be non-negative is not essential. We only require it to be bounded from below, and this is automatically assured by the compactness condition.

An example of theoretical interest. Let $\{P_\omega\}$ be the set of all probability distributions on the positive integers, E the set of even integers, $A = \{0, 1\}$, $L(\omega, 0) = 0$ if $P_\omega(E) > 1/2$ and $= 1$ if $P_\omega(E) \leqq 1/2$, $L(\omega, 1) = 1$ if $P_\omega(E) > 1/2$ and $= 0$ if $P_\omega(E) \leqq 1/2$, an action to be taken after observing one positive integer.

Let φ be a strategy, where $\varphi(n)$ is the probability of taking action 0 (saying "even") when integer n is observed. It can be shown, but only after a complicated argument, that φ is a Bayes solution if and only if it does "the right thing" for at least one integer n, i.e., if $\varphi(2n) = 1$ or $\varphi(2n - 1) = 0$ for at least one integer n. Further, all these strategies are admissible, but the class of all admissible strategies is not known. A plausible conjecture seems to be that φ is admissible if and only if $\inf_n \min \{\varphi(2n - 1), 1 - \varphi(2n)\} = 0$. Thus we see that even in the simple case of one observation, the classes are extremely large.

Let us consider this problem in the light of the invariance principle. Let g be the particular kind of permutation of the integers which, for some n, permutes the first n even numbers and the first n odd numbers separately and leaves the tail untouched. The set $\mathcal{G}$ of all such permutations for all n clearly forms a group, containing countably many elements g, which leaves the problem invariant. Let $\mathcal{C}$ be the set of all subsets of $\mathcal{G}$. The set Z of all integral outcomes breaks up into two orbits under the group, which we may write as E and O for "even" and "odd." The orbits Ω_s of Ω (closed in a suitable topology, a matter we need not go into here) are given by

$$\Omega_s = \{\omega : P_\omega(E) = s\}$$

for all s belonging to the closed interval [0, 1]. There is an asymptotically (right) invariant sequence of probability measures on $(\mathcal{G}, \mathcal{C})$, as follows: Let μ_n be the

measure which assigns equal probabilities of $1/(n!)^2$ to each of the $(n!)^2$ elements g affecting the first $2n$ integers only, and 0 to the remaining elements of $\mathcal{G}$. To prove the asymptotic invariance of this sequence, we simply note that given $g \in \mathcal{G}$, there is an integer n_0 such that g affects the first $2n_0$ integers only, so that for every $C \in \mathcal{C}$ we necessarily have $\mu_n(Cg) = \mu_n(C)$ for all $n \geqq n_0$. By the generalized Hunt-Stein theorem, then, the problem reduces to the classical binomial case, say of tossing a coin once with unknown probability s of showing heads (even).

5. Mixed games, and the use of previous experience. In this section we present a purely game-theoretic model for the modified minimax principle, and suggest how by means of this model and our previous experience the principle might be applied in practice to bridge the gap between the Bayes and minimax approach.

Mixed Games. Let there be given a pair of games $G_1 = (X_1, Y, M_1)$ and $G_2 = (X_2, Y, M_2)$ with the same Y-space of randomized strategies available to the second player. Let p and $q = 1 - p$ be fixed probabilities with which player I is to play the respective games. We shall call such a setup a mixed game for player I. A strategy for the first player is then equivalent to an ordered couple (x_1, x_2), $x_1 \in X_1$, $x_2 \in X_2$, signifying his choice of strategy depending on the game he must eventually play (say, after a coin with probability p of heads is tossed). The payoff function M is clearly given by

$$M((x_1, x_2), y) = pM_1(x_1, y) + qM_2(x_2, y).$$

We denote the mixed game by $G = (X_1 \times X_2, Y, M)$.

More generally, given $G_i = (X_i, Y, M_i)$ $i = 1, 2, \cdots$ and probabilities p_i of playing the game $G_i (p_i \geqq 0, \sum_i p_i = 1)$, the mixed game for player I is equivalent to the game $G = (X, Y, M)$ where

$$X = X_1 \times X_2 \times \cdots, x = (x_1, x_2, \cdots) \in X$$

is a strategy for the first player, and $M(x, y) = \sum_i p_i M_i(x_i, y)$.

In the most general case, we have games $G_s = (X_s, Y, M_s)$ with $s \in S$ an index set, $\mathcal{S}$ a Borel field in S, and P a probability measure on $\mathcal{S}$. A strategy for player I is now a function f on S with $f(s) \in X_s$ for each $s \in S$ (i.e., f is a point in the function space $X = \times_{s \in S} X_s$) and

$$M(f, y) = \int_S M_s(f(s), y)\, dP(s).$$

Let us now consider the minimax principle from the point of view of the second player. For simplicity we confine ourselves to the discrete case, although the argument is general. For each strategy $y \in Y$, the second player seeks to minimize $\sup_x M(x, y)$ over y. But $\sup_x M(x, y) = \sum_i p_i \sup_{x_i \in X_i} M_i(x_i, y) = \sum_i p_i \alpha(i, y)$ is just the average with respect to the p_i of the sliced up risk function in the modified minimax sense, where each slice is the old payoff function M_i for the ith game. (The point $(\alpha(1, y), \alpha(2, y), \cdots)$ is the generalization of the

α-β set for testing composite hypotheses.) Hence playing minimax in the mixed game is equivalent to playing Bayes in the sliced up game.

The use of previous experience as a slicing principle: In a paper on the use of previous experience in reaching statistical decisions, Hodges and Lehmann [8] propose a guess at an a priori distribution λ for the states of nature on the basis of previous experience, and a safeguard in case the guess is incorrect. If c is the minimax risk, they seek to minimize $\int \rho(\omega, \varphi)\, d\lambda(\omega)$ among all procedures φ satisfying the restriction $\sup_\omega \rho(\omega, \varphi) \leqq c + k$ where k is a given positive number. They call such a minimizing procedure φ_0 a restricted Bayes solution (we might say the Bayes principle is given a minimax restriction). If λ is suspect, k is made small; as confidence in λ grows, k may be raised. Note that as $k \to \infty$ or is large enough, the restricted Bayes solution approaches the Bayes solution; as $k \to 0$, the restricted solution approaches the minimax state (for $k = 0$, we have a best minimax procedure relative to λ).

Another way of looking at the modified minimax principle, with a view to applying it in practice, suggests that we reverse this procedure. That is to say, to start with the minimax principle and then modify it gradually or greatly as previous experience is gathered—to assume less and trust it completely, a sort of Bayesian modification of the minimax principle. Specifically, to take the most simple case, we might use the best of our information to break up or stratify Ω into Ω_1 and Ω_2 with certain probabilities p and $1 - p$ attached thereto, and then play minimax in the resulting mixed game. (Say, a machine is known to produce with probability $p = .95$ coins whose probabilities of showing heads when tossed lie between .4 and .6. Ω_1 is then the interval from .4 to .6, Ω_2 its complement in the unit interval, and the corresponding games are mixed in the ratio .95 to .05. If the probability that $\omega \varepsilon \Omega_1$ is at least p, then we would mix Ω_1 and all of Ω in the proportion p to $1 - p$, and so on.) As more knowledge is acquired, we might adjust the value of p, or better still, feel confident enough to partition Ω into a larger number of sets with certain p_i attached, or even into a family $\{\Omega_\alpha\}$ of sets with a probability measure P over it, and then play minimax in the resulting mixed game. By the above, this is equivalent to playing Bayes in the sliced up game. If each Ω_α consists of only one point, we are really postulating an a priori probability distribution P over the states of nature, so that minimax in the mixed game is exactly Bayes for the original game. Thus we see how previous information might be used in any given problem to slice Ω into subsets, and how it is possible to go from the minimax extreme to the Bayes in easy stages, by a gradual modification of the minimax principle.

Acknowledgment. I wish to express my deep personal indebtedness to the late Professor M. A. Girshick for suggesting this topic, and for his interest and guidance throughout my statistical studies. I am indebted to Professor Herman Rubin for several helpful discussions. I wish to thank the referee for pointing out references [9] and [10], as papers containing further useful remarks on the Hunt-Stein theorem and the role of invariance in Statistics; I am glad to call them to

the attention of the reader. To the Office of Naval Research my thanks for its support of this work.

REFERENCES

[1] David Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley & Sons, New York, 1954.
[2] Abraham Wald, *Statistical Decision Functions*, John Wiley & Sons, New York, 1950.
[3] Paul R. Halmos, *Measure Theory*, D. Van Nostrand, New York, 1950.
[4] G. Hunt and C. Stein, "Most stringent tests of composite hypotheses," unpublished.
[5] S. Banach, *Théorie des opérations linéaires*, Warszawa, 1932.
[6] Paul R. Halmos and L. J. Savage, "Application of the Radon-Nikodym theorem to the theory of sufficient statistics," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 225–241.
[7] J. L. Kelley, *General Topology*, D. Van Nostrand, New York, 1955.
[8] J. L. Hodges, Jr. and E. L. Lehmann, "The use of previous experience in reaching statistical decisions," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 396–407.
[9] E. L. Lehmann, "Some principles of the theory of testing hypotheses," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 1–26.
[10] J. Kiefer, "Invariance, minimax sequential estimation, and continuous time processes," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 573–601.

ON LINEAR ASSOCIATIVE ALGEBRAS CORRESPONDING TO ASSOCIATION SCHEMES OF PARTIALLY BALANCED DESIGNS

By R. C. Bose and Dale M. Mesner[1]

University of North Carolina and *Michigan State University*

1. Introduction. Given v objects $1, 2, \cdots, v$, a relation satisfying the following conditions is said to be an association scheme with m classes:

(a) Any two objects are either 1st, or 2nd, $\cdots$, or mth associates, the relation of association being symmetrical, i.e., if the object α is the ith associate of the object β, then β is the ith associate of α.

(b) Each object α has n_i ith associates, the number n_i being independent of α.

(c) If any two objects α and β are ith associates, then the number of objects which are jth associates of α, and kth associates of β, is p_{jk}^i and is independent of the pair of ith associates α and β.

The numbers v, n_i $(i = 1, 2, \cdots, m)$ and p_{jk}^i $(i, j, k = 1, 2, \cdots, m)$ are the parameters of the association scheme.

If we have an association scheme with m classes and given parameters, then we get a partially balanced design with r replications and b blocks if we can arrange the v objects into b sets (each set corresponding to a block) such that

(i) each set contains k objects (all different);

(ii) each object is contained in r sets;

(iii) if two objects α and β are ith associates, then they occur together in λ_i sets, the number λ_i being independent of the particular pair of ith associates α and β.

Partially balanced designs were introduced in experimental studies by Bose and Nair [5], and have recently come into fairly general practical use. The concept of the association scheme, though inherent in Bose and Nair's definition, was explicitly introduced by Bose and Shimamoto [6], as an aid to the classification and analysis of partially balanced designs.

2. Association schemes as concordant graphs. An association scheme with v objects and m classes may be visualized as follows:

Let the objects be points. Suppose we have m colors $C_1, C_2, \cdots, C_m$. If two objects are ith associates we connect them by a segment of the ith color. The points together with the segments of the ith color form a linear graph which will

Received February 21, 1958.

[1] The research by the first author was supported in part by the United States Air Force under contract AF 18(600)-83, monitored by the Office of Scientific Research.

Some of the work of the second author appeared in his doctoral dissertation at the Michigan State University [15]. He is now an NRC-NBS Post Doctoral Research Associate at the Statistical Engineering Laboratory, National Bureau of Standards.

be regular of degree n_i as a result of property (b). We may say that the n graphs together are *concordant*[2] when properties (a) and (c) are also satisfied, the meanings of these being as follows:

a) Every pair of points is connected by a single segment of one of the m colors. The graphs are non-oriented.

b) If any two points α and β are connected by a segment of the ith color, then the number of points which are connected to α by a segment of color C_j and to β by a segment of color C_k, is p_{jk}^i and is independent of the particular pair of points chosen.

Equivalently p_{jk}^i is the number of 2-chains directed from α to β and consisting of segments of colors C_j and C_k in that order. Clearly the p_{jk}^i are closely related to the number of triangles in the graph formed of segments of colors C_i, C_j, C_k. Properties (a), (b) and (c) are just enough to specify the number of segments of each color on each point, and the number of triangles of each combination of colors on each segment. The total number of segments, the total number of 2-chains, and the total number of triangles in the graph are then readily determined. Methods based on the incidence matrices of the graphs [16] can be used with (3.6) to enumerate certain chains of more than two segments. The arrangement in these graphs of all configurations involving two points or three points shows a striking regularity which does not extend to configurations having more than three points. It can be shown by examples that the points of the graph of color C_i may not all lie on the same number of complete 4-points, and that two association schemes with the same parameter values may give graphs differing in the total number of complete 4-points. This shows that the structure of concordant graphs is not determined completely by properties (a) to (c).

3. Association matrices. We define[3]

$$B_i = (b_{\alpha\beta}^i) = \begin{bmatrix} b_{11}^i & b_{12}^i & \cdots & b_{1v}^i \\ \cdots & \cdots & \cdots & \cdots \\ b_{v1}^i & b_{v2}^i & \cdots & b_{vv}^i \end{bmatrix},$$

where

$b_{\alpha\beta}^i = 1$, if the objects α and β are ith associates (or connected by a segment of the ith graph);

$= 0$, otherwise.

B_i is a symmetric matrix, in which each row total and each column total is n_i.

Let each object be the zeroth associate of itself and of no other treatment.

---

[2] Not to be confused with *chromatic* graphs, in which points, not segments, are colored. For a general discussion of linear graphs, see [11].

[3] The convention will be adopted here of using a superscript as the column index of a matrix, the first subscript as the row index, and the second subscript as the index of the matrix itself. This choice is dictated by the notation already established for the parameters p_{jk}^i.

Then

$$\begin{aligned} B_0 &= I_v, && \text{the } v \times v \text{ identity matrix,} \\ n_0 &= 1, \\ p_{ij}^0 &= n_i, && \text{if } i = j, \\ &= 0, && \text{otherwise,} \\ p_{0k}^i &= 1, && \text{if } i = k, \\ &= 0, && \text{otherwise,} \\ \lambda_0 &= r, && \text{for designs.} \end{aligned}$$

The following identities are known [5] and can be proved easily by combinatorial methods. Proofs based on the matrices B_i will be given in Section 5.

$$\begin{gathered} \sum_{i=0}^{m} n_i = v, \\ \sum_{j=0}^{m} p_{jk}^i = n_k, \\ p_{jk}^i = p_{kj}^i, \\ n_i p_{jk}^i = n_j p_{ik}^j = n_k p_{ij}^k. \end{gathered} \tag{3.1}$$

Further the following two identities hold for designs:

$$\begin{gathered} bk = vr, \\ \sum_{i=0}^{m} n_i \lambda_i = rk. \end{gathered} \tag{3.2}$$

Among the numbers

$$b_{\alpha 0}^{\beta}, \quad b_{\alpha 1}^{\beta}, \cdots, \quad b_{\alpha m}^{\beta}$$

only one is unity, i.e., $b_{\alpha i}^{\beta}$ if α and β are ith associates. Hence

$$B_0 + B_1 + \cdots + B_m = J_v, \tag{3.3}$$

where J_v is the $v \times v$ matrix each of whose elements is unity.

It also follows that the linear form

$$c_0 B_0 + c_1 B_1 + \cdots + c_m B_m \tag{3.4}$$

is equal to the zero matrix if and only if

$$c_0 = c_1 = \cdots = c_m = 0;$$

hence the linear functions of $B_0, B_1, \cdots, B_m$ form a vector space with basis $B_0, B_1, \cdots, B_m$.

LEMMA 3.1.

$$\sum_{\gamma=1}^{v} b_{\alpha j}^{\gamma} b_{\gamma k}^{\beta} = p_{jk}^{0} b_{\alpha 0}^{\beta} + \cdots + p_{jk}^{i} b_{\alpha i}^{\beta} + \cdots + p_{jk}^{m} b_{\alpha m}^{\beta}. \tag{3.5}$$

The objects α and β are zeroth associates if $\alpha = \beta$; otherwise they are either 1st, 2nd, $\cdots$ or mth associates from condition (a) of Section 1. Suppose they are ith associates. Both terms of the product $b_{\alpha j}^{\gamma} b_{\gamma k}^{\beta}$ are unity if and only if γ is the jth associate of α as well as kth associate of β. Hence from condition (c) Section 1, the left-hand side of (3.5) is p_{jk}^{i}. Again since α and β are ith associates $b_{\alpha l}^{\beta}$ is unity if $l = i$ and is zero otherwise. Hence the right-hand side of (3.5) is also equal to p_{jk}^{i}. This proves the Lemma.

We now note that the left-hand side of (3.5) is the element in the αth row and βth column of the product $B_j B_k$, and $b_{\alpha l}^{\beta}$ is the element in the αth row and βth column of B_l ($l = 0, 1, \cdots, m$). Thus[4]

$$B_j B_k = p_{jk}^{0} B_0 + p_{jk}^{1} B_1 + \cdots + p_{jk}^{m} B_m. \tag{3.6}$$

The product of two matrices of the form (3.4), where the c_i are scalars, may be expressed as a linear combination of terms of the form $B_j B_k$ and will reduce to the form (3.4). The set of matrices of this form is therefore closed under multiplication. It is clear that it forms an Abelian group under addition. Thus the linear functions of $B_0, B_1, \cdots, B_m$ form a ring with unit element, which will be a linear associative algebra if the coefficients c_i range over a field. Multiplication is also commutative. This statement and the equivalent statement $p_{jk}^{i} = p_{kj}^{i}$ will be shown in Section 5 to follow from (3.6) and the symmetry of B_i.

Linear associative algebras have of course been extensively studied and are treated, for example, in [13]. The properties of most importance in the present study are easily established, and brief proofs will now be given for the sake of completeness.

We first find the consequences of the associative law of matrix multiplication:

$$\begin{aligned} B_i(B_j B_k) &= B_i \sum_u p_{jk}^{u} B_u \\ &= \sum_u p_{jk}^{u} B_i B_u \\ &= \sum_{u,t} p_{jk}^{u} p_{iu}^{t} B_t. \end{aligned}$$

Also

$$\begin{aligned} (B_i B_j) B_k &= \Big(\sum_u p_{ij}^{u} B_u\Big) B_k \\ &= \sum_u p_{ij}^{u} B_u B_k \\ &= \sum_{u,t} p_{ij}^{u} p_{uk}^{t} B_t. \end{aligned}$$

[4] The fundamental formula (3.6) was first obtained by W. A. Thompson [17], [19] and was independently discovered by the second author [15]. Other results of Section 3 were included in a set of lectures [2] at the University of Frankfurt by the first author. Some of these were independently obtained in another form by the second author. When the two authors learned of each other's work, they decided to collaborate in a joint paper.

From the independence of $B_0, B_1, \cdots, B_m$,

$$\sum_u p_{ij}^u p_{uk}^t = \sum_u p_{jk}^u p_{iu}^t. \tag{3.7}$$

In these equations the summation over u runs from 0 to m and the remaining indices are arbitrary but fixed,

$$0 \leqq i, j, k, t \leqq m$$

Now let us define $\mathcal{P}_k$ by[5]

$$\mathcal{P}_k = (p_{ik}^j) = \begin{bmatrix} p_{0k}^0 & p_{0k}^1 & \cdots & p_{0k}^m \\ p_{1k}^0 & p_{1k}^1 & \cdots & p_{1k}^m \\ \cdots & \cdots & \cdots & \cdots \\ p_{mk}^0 & p_{mk}^1 & \cdots & p_{mk}^m \end{bmatrix}, \qquad k = 1, 0, \cdots, m.$$

Now the left side of (3.7) is the element in the ith row and tth column of $\mathcal{P}_j\mathcal{P}_k$. Also the element in the ith row and tth column of $\mathcal{P}_u$ is p_{iu}^t, so that the right side of (3.7) is the element in the ith row and tth column of

$$p_{jk}^0\mathcal{P}_0 + p_{jk}^1\mathcal{P}_1 + \cdots + p_{jk}^m\mathcal{P}_m.$$

Hence we have

$$\mathcal{P}_j\mathcal{P}_k = p_{jk}^0\mathcal{P}_0 + p_{jk}^1\mathcal{P}_1 + \cdots + p_{jk}^m\mathcal{P}_m. \tag{3.8}$$

Thus, the $\mathcal{P}$'s multiply in the same manner as the B's. Since $p_{0k}^i = 1$ if $k = i$ and 0 otherwise, the 0th row of $\mathcal{P}_k$ contains a 1 in column k and 0's in other positions, which is enough to show that if

$$c_0\mathcal{P}_0 + c_1\mathcal{P}_1 + \cdots + c_m\mathcal{P}_m = 0,$$

then

$$c_0 = c_1 = \cdots = c_m = 0;$$

i.e., $\mathcal{P}_0, \mathcal{P}_1, \cdots, \mathcal{P}_m$ are linearly independent. They thus form the basis for a vector space and combine in the same way as the B's under addition, as well as under multiplication. They provide a regular representation in

$$(m + 1) \times (m + 1)$$

matrices of the algebra given by the B's, which are $v \times v$ matrices. In particular, $\mathcal{P}_0 = I_{m+1}$.

Since the B's are commutative, the $\mathcal{P}$'s are commutative. In general they are not incidence matrices and are not symmetric. $\mathcal{P}_k$ does not have equal row totals, but has the same equal column totals n_k as B_k. In analogy with (3.3), all elements of row j of $\sum_k \mathcal{P}_k$ are equal to n_j. Let

$$B = c_0B_0 + c_1B_1 + \cdots + c_mB_m$$

[5] It should be noted that these matrices differ from matrices $P_k = (p_{ij}^k)$ which were defined in several earlier papers ([4], [5], [6]) but do not have the same algebraic properties.

be any element of our algebra, and let $f(\lambda)$ be a polynomial. Then we can express

$$f(B) = l_0B_0 + l_1B_1 + \cdots + l_mB_m .$$

If

$$\mathcal{P} = c_0\mathcal{P}_0 + c_1\mathcal{P}_1 + \cdots + c_m\mathcal{P}_m$$

is the representation of B, then

$$f(\mathcal{P}) = l_0\mathcal{P}_0 + l_1\mathcal{P}_1 + \cdots + l_m\mathcal{P}_m .$$

Let $f(\lambda)$ be the minimum function of B and $\phi(\lambda)$ the minimum function of $\mathcal{P}$. Then $f(\lambda)$ is the monic polynomial of least degree for which

$$f(B) = 0.$$

$$f(B) = 0 \rightarrow l_0 = l_1 = \cdots = l_m = 0 \rightarrow f(\mathcal{P}) = 0;$$

i.e., $f(\lambda)$ is divisible by $\phi(\lambda)$.

Similarly $\phi(\lambda)$ is divisible by $f(\lambda)$. Since both are monic polynomials,

$$f(\lambda) = \phi(\lambda).$$

That is, B and $\mathcal{P}$ have the same distinct characteristic roots, and every matrix B has at most $m + 1$ distinct characteristic roots, which are solutions of the minimum equation of $\mathcal{P}$.

4. Applications to combinatorial problems. Association matrices will be used to derive some results first obtained in [9] by a longer method.

The incidence matrix $N = (n_{ij})$ of a design is defined by

$$n_{ij} = 1, \quad \text{if treatment } i \text{ occurs in block } j,$$
$$= 0, \quad \text{otherwise.}$$

Then

$$B = NN' = rB_0 + \lambda_1B_1 + \cdots + \lambda_mB_m ,$$
$$\mathcal{P} = r\mathcal{P}_0 + \lambda_1\mathcal{P}_1 + \cdots + \lambda_m\mathcal{P}_m .$$

Also

$$C = r\left(1 - \frac{1}{k}\right)B_0 - \frac{\lambda_1}{k}B_1 - \cdots - \frac{\lambda_m}{k}B_m ,$$

where C is the coefficient matrix in the normal equations for estimating the treatment effects after adjusting for the block effects [6]. Clearly C is a symmetric matrix. If e is a characteristic root of C, then $k(r - e)$ is a characteristic root of B. It is known that C has rank $v - 1$ for a connected design [1]. In this case,[6] therefore, 0 is a simple root of C and rk is a simple root of B, a fact which could also be shown directly as follows.

[6] Connectedness was assumed implicitly in [9].

The elements of B or NN' are non-negative, and for connected designs B is irreducible. Also it is easy to verify that the sum of the elements in any row or column of B is $\sum n_i\lambda_i = rk$. Hence

$$B^* = \frac{1}{rk} B = \frac{1}{rk} NN'$$

is a stochastic matrix, which shows that unity is a simple root of B^* and is greater than all the other roots [7]. Thus rk is a simple root of B. The results of Section 3 show that rk is a root of $\mathcal{P}$ and exceeds the other roots. If this root is removed from $|\mathcal{P} - I\theta| = 0$, then for the case $m = 2$, the other two characteristic roots of $\mathcal{P}$ will be roots of a quadratic equation which reduces to

$$(r-\theta)^2 + [(\lambda_1 - \lambda_2)(p_{12}^2 - p_{12}^1) - (\lambda_1 + \lambda_2)](r-\theta) + [(\lambda_1 - \lambda_2)(\lambda_2 p_{12}^1 - \lambda_1 p_{12}^2) + \lambda_1\lambda_2] = 0,$$

on using the identities (3.1) and (3.2). The roots are given by

$$\begin{aligned} r - \theta_1 &= \tfrac{1}{2}[(\lambda_1 - \lambda_2)(-\gamma - \sqrt{\Delta}) + (\lambda_1 + \lambda_2)], \\ r - \theta_2 &= \tfrac{1}{2}[(\lambda_1 - \lambda_2)(-\gamma + \sqrt{\Delta}) + (\lambda_1 + \lambda_2)], \end{aligned} \tag{4.1}$$

where

$$\gamma = p_{12}^2 - p_{12}^1, \qquad \beta = p_{12}^1 + p_{12}^2, \qquad \Delta = \gamma^2 + 2\beta + 1.$$

Therefore,

$$|NN' - I_v\theta| = (rk - \theta)(\theta_1 - \theta)^{\alpha_1}(\theta_2 - \theta)^{\alpha_2}.$$

To determine the multiplicities α_1 and α_2 we note that

$$\begin{aligned} \operatorname{Tr} I_v &= 1 + \alpha_1 + \alpha_2 = v, \\ \operatorname{Tr} NN' &= rk + \alpha_1\theta_1 + \alpha_2\theta_2 = vr. \end{aligned}$$

Solving and using (4.1),

$$\begin{aligned} \alpha_1 &= \frac{n_1 + n_2}{2} - \frac{(n_1 - n_2) + \gamma(n_1 + n_2)}{2\sqrt{\Delta}}, \\ \alpha_2 &= \frac{n_1 + n_2}{2} + \frac{(n_1 - n_2) + \gamma(n_1 + n_2)}{2\sqrt{\Delta}}. \end{aligned} \tag{4.2}$$

Thus the multiplicities α_1 and α_2 of the roots of NN' are determined in terms of the parameters of the design. It is striking that, being independent of r and λ_i, these multiplicities are the same for all designs having a given association scheme. This is an instance of some general properties of α_i to be established in Section 6.

For a design to exist, α_1 and α_2 must be integral. The condition this imposes on the parameters appearing in (4.2) has been used in studies of the existence and non-existence of designs [3], [8], [9], [12], [15].

5. Applications of algebraic properties of association matrices. In this section we assume only that B_i $(i = 0, \cdots, m)$, are symmetric incidence matrices satisfying

$$B_0 = I_v, \tag{5.1}$$

$$\sum_{i=0}^{m} B_i = J_v, \tag{5.2}$$

$$B_j B_k = \sum_{i=0}^{m} p_{jk}^i B_i, \tag{5.3}$$

for some set of constants p_{jk}^i. All of the properties of the algebra except commutativity follow immediately, including its representation in terms of the matrices $\mathcal{P}_k = (p_{ik}^j)$. Also, p_{jk}^i are elements of products of incidence matrices and must be non-negative integers. From

$$B_k = B_0 B_k = \sum_i p_{0k}^i B_i$$

we deduce the special values

$$p_{0k}^i = 1, \qquad \text{if } i = k,$$
$$= 0, \qquad \text{if } i \neq k.$$

The diagonal element in row t, column t of $B_j B_k'$ may be interpreted as the number of positions occupied by 1's in row t of B_j as well as in row t of B_k. (5.2) shows that if $k \neq j$ this element is zero. If $k = j$ it is equal to the number of 1's in row t of B_j. The expansions of $B_j B_k' = B_j B_k$ and $B_j B_j' = B_j^2$ then show that

$$p_{jk}^0 = 0, \qquad j \neq k,$$

and that p_{jj}^0 is equal to each row total of B_j. These row totals must therefore be equal. As a matter of notation set

$$p_{jj}^0 = n_j.$$

Row totals in (5.2) show

$$\sum n_j = v.$$

Also

$$(\sum_j B_j) B_k = J_v B_k = n_k J_v = \sum_i n_k B_i,$$

and

$$\sum_j (B_j B_k) = \sum_j \sum_i p_{jk}^i B_i = \sum_i (\sum_j p_{jk}^i) B_i.$$

Hence comparing coefficients,

$$\sum_j p_{jk}^i = n_k.$$

We now show that commutativity follows from symmetry of B_i.

$$B_k B_j = B_k' B_j' = (B_j B_k)' = \Big(\sum_i p_{jk}^i B_i\Big)'$$

$$= \sum_i p_{jk}^i B_i' = \sum_i p_{jk}^i B_i = B_j B_k .$$

As a consequence,

$$p_{jk}^i = p_{kj}^i .$$

We also deduce

$$\mathcal{P}_j \mathcal{P}_k = \mathcal{P}_k \mathcal{P}_j .$$

Equating the elements in the sth row and tth column of $\mathcal{P}_j\mathcal{P}_k$ and $\mathcal{P}_k\mathcal{P}_j$,

$$\sum_i p_{sj}^i p_{ik}^t = \sum_i p_{sk}^i p_{ij}^t . \tag{5.4}$$

This relation is equivalent to (3.7). Taking $t = 0$ we get

$$\sum_i p_{sj}^i p_{ik}^0 = \sum_i p_{sk}^i p_{ij}^0 ,$$

$$n_k p_{sj}^k = n_j p_{sk}^j .$$

We have now shown that all the known identities (3.1) follow from the properties of the algebra which were stated at the beginning of this section. However the relation

$$\mathcal{P}_j \mathcal{P}_k = \mathcal{P}_k \mathcal{P}_j$$

leads to new identities when $m > 2$.

To prove a new identity in the case $m = 3$, set $j = s = 1$, $k = t = 2$ in (5.4), giving

$$n_1 + p_{11}^1 p_{12}^2 + p_{11}^2 p_{22}^2 + p_{11}^3 p_{32}^2 = p_{12}^1 p_{11}^2 + p_{12}^2 p_{21}^2 + p_{12}^3 p_{31}^2 .$$

We remark that when $m = 3$, other choices of j, k, s, t lead to relations equivalent to this one. The use of a smaller number of parameters will make it easier to recognize equivalent expressions and will be helpful in simplifying the identity. A fairly symmetric set of parameters is the following:

$$\begin{aligned} &n_1 , \quad n_2 , \quad n_3 , \\ a_{12} &= n_1 p_{22}^1 = n_2 p_{12}^2 , \\ a_{31} &= n_3 p_{11}^3 = n_1 p_{13}^1 , \\ a_{23} &= n_2 p_{33}^2 = n_3 p_{23}^3 , \\ x &= n_1 p_{23}^1 = n_2 p_{13}^2 = n_3 p_{12}^3 . \end{aligned}$$

Known identities can be used to express all p_{jk}^i in terms of these parameters, whereupon the above identity reduces to

$$x^2\left(\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}\right)+x\left(\frac{a_{12}}{n_1}+\frac{a_{22}}{n_2}+\frac{a_{31}}{n_3}-n_1-n_2-n_3-1\right)$$
$$+\left(\frac{a_{31}\,a_{12}}{n_1}+\frac{a_{12}\,a_{22}}{n_2}+\frac{a_{22}\,a_{31}}{n_3}-n_1\,a_{23}-n_2\,a_{31}-n_3\,a_{12}+n_1\,n_2\,n_3\right)=0. \tag{5.5}$$

Thus when n_1, n_2, n_3, a_{12}, a_{23}, a_{31} are given, x must satisfy a quadratic equation. This is a new relation, since known identities (3.1) do not determine x in terms of the other chosen parameters. An example will illustrate this. Let

$$n_i = 8, \qquad a_{ij} = 24.$$

Then sets of p_{jk}^i which satisfy (3.1) are obtained for

$$x = 8,\ 16,\ 24,\ 32 \text{ or } 40.$$

However, (5.5) becomes

$$3x^2/8 - 16x + 152 = 0$$

and has no integral solutions, showing that the parameter values $n_i = 8$, $a_{ij} = 24$ are impossible.

An equivalent and perhaps easier way to impose the new necessary conditions on a given set of parameters is to form the matrix products $\mathcal{P}_j\mathcal{P}_k$ and $\mathcal{P}_k\mathcal{P}_j$ and require that they be identical.

The property of symmetry in the matrices B_i was used in the proof of commutativity in the algebra, which has been of key importance in the proofs of several of the foregoing identities. The fact that the elements of B_i are 0's and 1's has been used in determining the special form of the p_{jk}^i values but has not been vital in the algebra or the identities relating p_{jk}^i. The simple example

$$B_0=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix},\qquad B_1=\begin{bmatrix}0&1&0\\1&-1&1\\0&1&0\end{bmatrix},\qquad B_2=\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix}$$

shows that matrices with elements other than 0's and 1's may have the same algebraic behavior as association matrices and may lead to the same identities. This shows the necessity of the word "incidence" in the following lemma, which summarizes several results of this section.

LEMMA 5.1. *If B_i, $i = 0, 1, \cdots, m$ are symmetric incidence matrices satisfying*

$$B_0 = I_v,$$

$$\sum_{i=0}^{v} B_i = J_v,$$

$$B_j B_k = \sum_{i=0}^{m} p_{jk}^i B_i$$

for some set of constants p_{jk}^i, then B_i are the association matrices of an association scheme satisfying (a) *to* (c) *of Section* 1.

This lemma provides a useful algebraic method of verifying whether a given association relation satisfies the conditions of partial balance.

Algebraic sufficiency conditions may be used for designs as well as association schemes. It is easy to verify that an incidence matrix N is the matrix of a PBIB design if and only if N has equal column totals and

$$NN' = rB_0 + \lambda_1 B_1 + \cdots + \lambda_m B_m$$

for some m and some numbers $r, \lambda_1, \cdots, \lambda_m$, where $B_0, B_1, \cdots, B_m$ satisfy the conditions of Lemma 5.1. An application of this Lemma will be made in the proof of the next theorem.

Given an association scheme $\mathfrak{A}$ with more than m classes, let the indices of the associate classes be arranged into disjoint sets $S_0 = \{0\}, S_1, \cdots, S_m$. Define a new association relation $\mathfrak{B}$ in which associate classes correspond to sets S_i, two treatments being defined as ith associates in $\mathfrak{B}$ if and only if the associate class of the two treatments in $\mathfrak{A}$ corresponds to one of the indices in set S_i.

Association relations obtained in this way do not in general satisfy the conditions of partial balance. Lemma 4.1 of [18] states necessary and sufficient conditions for partial balance in the case $S_1 = \{1, 2\}, S_i = \{i + 1\}, i \geqq 2$, i.e., the case in which just two classes are combined. Iteration may give schemes in which several classes have been combined. However, examples are known [15] in which a combination of 3 or more classes will give a new scheme with partial balance while every combination of 2 classes fails, so that the iterative procedure is impossible. The following generalization is therefore nontrivial.

THEOREM 5.1. *Given an association scheme $\mathfrak{A}$ with v treatments and parameter values $q^{\alpha}_{\beta\gamma}$, let an association relation $\mathfrak{B}$ with v treatments have classes $0, 1, \cdots, m$ determined by disjoint sets $S_0 = \{0\}, S_1, \cdots, S_m$ of indices of $\mathfrak{A}$. In order for $\mathfrak{B}$ to satisfy the conditions of partial balance it is n.a.s. that there exist constants p^i_{jk} such that*

$$\sum_{\beta \in S_j} \sum_{\gamma \in S_k} q^{\alpha}_{\beta\gamma} = p^i_{jk} \tag{5.6}$$

uniformly for $\alpha \in S_i$, and for $i, j, k = 0, 1, \cdots, m$; in this case $\mathfrak{B}$ has parameter values p^i_{jk}.

PROOF. We denote incidence matrices of $\mathfrak{A}$ by A_α and of $\mathfrak{B}$ by B_i. From the definition of $\mathfrak{B}$,

$$B_i = \sum_{\alpha \in S_i} A_\alpha .$$

Lemma 5.1 will now be applied.

Clearly B_i are symmetric incidence matrices, $B_0 = I_v$ and $\sum_i B_i = J_v$; in order for $\mathfrak{B}$ to have partial balance it is thus n.a.s. that there exist constants p^i_{jk} such that

$$B_j B_k = \sum_{i=0}^{m} p^i_{jk} B_i .$$

Substituting,

$$B_j B_k = \Big(\sum_{\beta \in S_j} A_\beta\Big)\Big(\sum_{\gamma \in S_k} A_\gamma\Big)$$
$$= \sum_{\beta \in S_j} \sum_{\gamma \in S_j} A_\beta A_\gamma = \sum_{\beta \in S_j} \sum_{\gamma \in S_k} \sum_{\alpha} q^{\alpha}_{\beta\gamma} A_\alpha$$
$$= \sum_{\alpha} \Big(\sum_{\beta \in S_j} \sum_{\gamma \in S_k} q^{\alpha}_{\beta\gamma}\Big) A_\alpha .$$

Also

$$\sum_{i=0}^{m} p^i_{jk} B_i = \sum_{i=0}^{m} p^i_{jk} \sum_{\alpha \in S_i} A_\alpha$$
$$= \sum_{i=0}^{m} \sum_{\alpha \in S_i} p^i_{jk} A_\alpha ;$$

the coefficient of A_α in this expression has the same value p^i_{jk} for every $\alpha \in S_i$. Equating to the coefficient of A_α in the previous equation we obtain (5.6) as the n.a.s. conditions on the parameter values $q^{\alpha}_{\beta\gamma}$ of $\mathfrak{B}$, completing the proof of the theorem.

6. Characteristic roots of matrices in the algebras. The procedure used in Section 4 to determine the multiplicities α_1 and α_2 is readily generalized to association schemes with m classes. If θ is a characteristic root of $B = \sum_{i=0}^{m} c_i B_i$, where c_i belong to the field of real (or complex) numbers, then θ^n is a characteristic root of B^n. Also, the trace of any matrix is equal to the sum of its characteristic roots. This leads to a system of equations in the roots θ_u of $\mathcal{P} = \sum_{i=0}^{m} c_i \mathcal{P}_i$ and the multiplicities α_u of the same roots of B. θ_0 will designate $\sum_{i=0}^{m} c_i n_i$, the common value of the row totals of B.

$$\begin{aligned}
\alpha_0 + \alpha_1 + \cdots + \alpha_m &= \operatorname{Tr} I, \\
\alpha_0\theta_0 + \alpha_1\theta_1 + \cdots + \alpha_m\theta_m &= \operatorname{Tr} B, \\
\alpha_0\theta_0^2 + \alpha_1\theta_1^2 + \cdots + \alpha_m\theta_m^2 &= \operatorname{Tr} B^2, \qquad (6.1) \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
\alpha_0\theta_0^m + \alpha_1\theta_1^m + \cdots + \alpha_m\theta_m^m &= \operatorname{Tr} B^m.
\end{aligned}$$

Equations of this form were used in [9] but were limited to the cases $m \leqq 4$ because of the difficulty of computing $\operatorname{Tr} B^n$ with methods then available. (3.6) may be used to express B^n in the form

$$B^n = c_{0,n}B_0 + c_{1,n}B_1 + \cdots + c_{m,n}B_m .$$

Then, since B_0 is the only B_i with non-zero diagonal elements,

$$\operatorname{Tr} B^n = v c_{0,n} .$$

The right members of the equations are therefore easily computed. The coefficients of α_u form the Vandermonde matrix with determinant

$$\prod_{0 \leq j < k \leq m} (\theta_k - \theta_j).$$

The system will therefore have a unique solution if and only if the $m + 1$ roots θ_u are distinct. It will be shown in Corollary 6.2 that this will be the case for at least some choice of c_i.

The solutions α_i must be non-negative integers. If they can be expressed in terms of the parameters c_i and p_{jk}^i this requirement will provide necessary conditions which the parameters must satisfy in order for matrix B to exist. An explicit solution of (6.1) requires a general solution of the equation $|\theta I - \mathcal{P}| = 0$, which may be difficult to obtain for $m > 2$, but one observation may be made at once. If the basis matrices B_i exist, then matrix B will exist for arbitrary values of c_i, with characteristic roots which obviously occur with integral multiplicities. This indicates that the integral nature of α_u must be independent of c_i and dependent only on p_{jk}^i. Theorem 6.3 will show that this holds not only for the integral property but for the exact values α_u. This is somewhat surprising in view of the form of (6.1), since the values θ_u and $\operatorname{Tr} B^n$ depend strongly on c_i. The other theorems of this section will give further insight into the nature of the roots θ_u and multiplicities α_u, as well as simplifying their computation. Results related to some of these have been obtained independently and by a different approach in [10].

It was pointed out in Section 4 that $\alpha_0 = 1$ for the matrix NN' if it is irreducible (which is the case when the design is connected). The same theorems for stochastic matrices [7] apply to any B with non-negative coefficients c_i. In particular any matrix B_i which is irreducible has n_i as a simple root. It follows from theorems (6.1) through (6.3), which we now proceed to prove, that θ_0 is a simple root (i.e., $\alpha_0 = 1$) for any set of coefficients c_i for which

$$B = \sum_{i=0}^{m} c_i B_i$$

is irreducible.

THEOREM 6.1. *Let the characteristic roots of $\mathcal{P}_i$ be z_{ui}, $u = 0, 1, \cdots, m$. Then for a suitable ordering of z_{ui} for each i, the characteristic roots of the matrix*

$$\mathcal{P} = \sum_{i=0}^{m} c_i \mathcal{P}_i$$

are given by

$$\theta_u = \sum_{i=0}^{m} c_i z_{ui}, \qquad u = 0, 1, \cdots, m. \tag{6.2}$$

PROOF. The matrices $\mathcal{P}_0, \cdots, \mathcal{P}_m$ are pairwise commutative. Frobenius' Theorem ([14], Theorem 16.1) then states that for a suitable ordering of the characteristic roots of z_{ui} of each $\mathcal{P}_i$, and for any rational function

$$f(x_0, \cdots, x_m)$$

the roots of

$$f(\mathcal{P}_0, \cdots, \mathcal{P}_m)$$

are given by

$$f(z_{u0}, \cdots, z_{um}), \qquad u = 0, 1, \cdots, m.$$

Also, the ordering of the roots is the same for every rational function f. The required theorem follows by taking

$$f(x_0, \cdots, x_m) = \sum_{i=0}^{m} c_i x_i .$$

COROLLARY 6.1. *The distinct characteristic roots θ_u of*

$$B = \sum_{i=0}^{m} c_i B_i$$

are given by

$$\theta_u = \sum_{i=0}^{m} c_i z_{ui}, \qquad u = 0, 1, \cdots, m.$$

The problem of finding θ_u is therefore solved if the values z_{ui} can be found and ordered. When they are ordered as specified by Theorem 6.1, we define the matrix

$$Z = (z_{ui}).$$

Since z_{ui} are the characteristic roots of real symmetric matrices B_i, Z is a real matrix.

THEOREM 6.2. *$Z = (z_{ui})$ is non-singular.*

PROOF. Let

$$y_0, y_1, \cdots, y_m$$

be a real solution of the system of homogeneous equations

$$\sum_{i=0}^{m} z_{ui} y_i = 0, \qquad u = 0, 1, \cdots, m. \tag{6.3}$$

This system has coefficient matrix Z and will have a non-zero solution if and only if Z is singular. Since Z is real there is no loss of generality in taking y_i real. By Corollary 6.1 the characteristic roots of the matrix

$$B = \sum_{i=0}^{m} y_i B_i$$

are given by the left side of (6.3) and are therefore all equal to zero. The sum of all products of roots taken s at a time is thus equal to zero, $s = 1, 2, \cdots, v$; this sum is equal to the generalized trace $\mathrm{Tr}_s B$, the sum of all $s \times s$ principal minor determinants of B. B is symmetric with diagonal elements y_0 and other elements $y_1, \cdots, y_m$. This follows by noting that among the incidence matrices $B_0, B_1, \cdots, B_m$ there is one and only one, say B_i, for which the element in the tth row and uth column is unity, whereas for B_j, $j \neq i$, the corresponding ele-

ment is zero. Hence B will have y_i in positions which correspond to unities of B_i. In particular the diagonal elements of B will all be y_0. Therefore

$$\mathrm{Tr}_1 B = vy_0 = 0,$$

giving

$$y_0 = 0.$$

Any element y_i $(i = 1, 2, \cdots, m)$ contributes $y_0^2 - y_i^2$ or $-y_i^2$ to $\mathrm{Tr}_2 B$. Since each row of B_i sums up to n_i, the number of unities above the diagonal in B_i s $vn_i/2$. This is also the number of y_i's above the diagonal in B. Hence

$$\mathrm{Tr}_2 B = -\frac{v}{2}(n_1 y_1^2 + n_2 y_2^2 + \cdots + n_m y_m^2) = 0.$$

Since $v, n_1, \cdots, n_m$ are positive integers and $y_1, y_2, \cdots, y_m$ are real it follows that

$$y_1 = y_2 = \cdots = y_m = 0.$$

Therefore (6.3) has no non-zero solution and Z is non-singular.

COROLLARY 6.2. *Given a set of association matrices $B_0, B_1, \cdots, B_m$, any ordered set of real (or complex) numbers $\theta_0, \cdots, \theta_m$ is the ordered set of distinct characteristic roots of*

$$B = c_0 B_0 + c_1 B_1 + \cdots + c_m B_m$$

for a unique set of real (or complex) coefficients c_i. In particular, matrices B exist with $m + 1$ distinct roots.

PROOF. For arbitrary $\theta_0, \cdots, \theta_m$ the system (6.2) can be solved uniquely for $c_0, \cdots, c_m$.

THEOREM 6.3. *If*

$$B = c_0 B_0 + c_1 B_1 + \cdots + c_m B_m$$

is an element of an algebra with the association matrices B_i as basis, then

$$|\theta I - B| = \prod_{u=0}^{m} (\theta - \theta_u)^{\alpha_u},$$

where α_u are independent of $c_0, \cdots, c_m$.

PROOF. Let S be an element of the algebra which has $m + 1$ distinct characteristic roots. Then S does not satisfy any polynomial equation with degree less than $m + 1$.

$$\begin{aligned}
S_0 &= B_0 = I, \\
S &= b_{01}B_0 + b_{11}B_1 + \cdots + b_{s1}B_m, \\
S^2 &= b_{02}B_0 + b_{12}B_1 + \cdots + b_{m2}B_m, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
S^m &= b_{0m}B_0 + b_{1m}B_1 + \cdots + b_{mm}B_m.
\end{aligned}$$

Since S does not satisfy any equation of degree m or less, these equations must be independent and can be solved to give each B_i as a linear expression in S^j with constant coefficients. Hence any arbitrary element B can be written

$$B = d_0 I + d_1 S + \cdots + d_m S^m.$$

If ϕ is a characteristic root of S, then the corresponding characteristic root of B will be

$$\theta = d_0 + d_1\phi + \cdots + d_m\phi^m.$$

All of the roots θ_μ of B may be obtained in this way by using all of the roots of S. If a root ϕ_μ has multiplicity α_μ, then the corresponding α_μ roots of B will be equal. That is, the roots θ_μ of an arbitrary matrix B have the same multiplicities α_μ as the corresponding roots ϕ_μ of the fixed matrix S and are therefore independent of the coefficients c_i occurring in B.

This completes the proof but an additional remark should be made. The element B may be such that distinct roots ϕ lead to the same value $\hat{\theta}$, whose multiplicity $\hat{\alpha}$ will be equal to the sum of two or more α_μ. In general, if M is a subset of the set $0, 1, \cdots, m$, $\theta_\mu = \hat{\theta}$ for $\mu \varepsilon M$, and $\theta_\mu \neq \hat{\theta}$ for $\mu \varepsilon M$, then

$$\hat{\alpha} = \sum_{\mu \varepsilon M} \alpha_\mu .$$

The statement of the theorem is correct whether θ_μ are distinct or not.

If the roots z_{ui} are obtained separately for each $\mathcal{P}_i$, it may not be immediately clear what ordering of them is required by Theorem 6.1. However, each z_{ui} is a root of B_i with multiplicity α_u. If the multiplicities are known, a suitable ordering will then be determined by any ordering of the α_u if the α_u are distinct, and partially determined if they are not all distinct. Theorem 6.5 will give another technique for ordering the roots. Theorem 6.4 reveals another significance of the distinctness or equality of the α_u.

THEOREM 6.4. *If t and only t of the multiplicities α_u are equal, then for each i the corresponding roots z_{ui} satisfy a monic polynomial equation with integral coefficients and degree t. In particular, if any α_u is distinct from the other multiplicities, the corresponding roots z_{ui} are rational integers.*

PROOF. The term m-polynomial will denote "monic polynomial with integral coefficients." The characteristic polynomial of a matrix with integral coefficients is an m-polynomial. Denote the characteristic polynomials of a basis matrix B_i and its representation $\mathcal{P}_i$ by

$$f_i(\theta) = |\theta I - B_i| = \prod_{u=0}^{m} (\theta - z_{ui})^{\alpha_u},$$

$$\phi_i(\theta) = |\theta I - \mathcal{P}_i| = \prod_{u=0}^{m} (\theta - z_{ui}).$$

For a particular root z_{ui}, let $g(\theta)$ be the m-polynomial of lowest degree s with z_{ui} as a zero. $g(\theta)$ is irreducible over the rational field. It is determined uniquely by any of its zeros and any m-polynomial which has any of its zeros is divisible by $g(\theta)$ and has all of its zeros ([13], Sec. 38). Therefore $f_i(\theta)$ and $\phi_i(\theta)$ are divisible

by $g(\theta)$, which must be the product of s of the linear factors of $\phi_i(\theta)$. Moreover, the corresponding multiplicities must all have the same value $\alpha_{u'}$; otherwise, after a certain number of successive divisions of $f_i(\theta)$ by $g(\theta)$ the quotient would be an m-polynomial which has some of the zeros of $g(\theta)$ but not all. In short, $f_i(\theta)$ contains $[g(\theta)]^{\alpha_{u'}}$ as a factor and at least s of the multiplicities are equal. It may happen that the set of distinct irreducible factors with multiplicity $\alpha_{u'}$ includes others along with $g(\theta)$. The product of the factors in the set will be the polynomial of degree t described in this theorem, where t is the sum of the degree s of $g(\theta)$ and the degrees of any other factors in the set. Clearly $s \leqq t$. If $t = 1$, then $s = 1$ and $g(\theta) = \theta - z_{u'i}$. Since $g(\theta)$ has integral coefficients, it follows in this case that $z_{u'i}$ is an integer, $i = 0, \cdots, m$.

Theorems 6.1 and 6.4 are illustrated in the case of $m = 2$ associate classes by expressions (4.1) for the roots θ_1 and θ_2, the roles of c_0, c_1, c_2 being played by r, λ_1, λ_2. Although in general the roots of a quadratic equation are irrational functions of the coefficients and although λ_1 and λ_2 occur several times in the coefficients, the roots in this case are linear polynomials in r, λ_1, λ_2, with coefficients that are rational if and only if the integer Δ is a perfect square. It is shown in [9] that if $\alpha_1 \neq \alpha_2$ it is in fact necessary that Δ be a perfect square, implying that the roots are rational. The additional fact that they are integers is not obvious from (4.1). It is further shown in [9] that if $\alpha_1 = \alpha_2$, it is possible that Δ will not be a square and that the roots will be irrational. This is precisely the case in known designs of cyclic type.

THEOREM 6.5. *For fixed $u = 0, \cdots, m$, the roots z_{ui} satisfy the relations*

$$z_{uj} z_{uk} = \sum_{i=0}^{m} p_{jk}^{i} z_{ui}. \tag{6.4}$$

PROOF. The relation is proved by applying Frobenius' theorem to both sides of (3.8).

It is interesting to note the amount of simplification that has now been made in the study of a matrix of the algebra, for example the matrix

$$NN' = rB_0 + \lambda_1 B_1 + \cdots + \lambda_m B_m$$

of a design. The characteristic equation of this matrix is of degree v. The regular representation introduced in Section 3 reduces its solution to the solution of an equation of degree $m + 1$. The theorems of this section show that the characteristic roots are linear combinations of $r, \lambda_1, \cdots, \lambda_m$ and that the multiplicities are entirely independent of these parameters, depending only on the association scheme. The coefficients of $r, \lambda_1, \cdots, \lambda_m$ are z_{ui}, the characteristic roots of the matrices $\mathcal{P}_i$, which also depend only on the association schemes. In some cases the z_{ui} can be shown to be integers and in any case they satisfy the system of quadratic equations (6.4). Once z_{ui} values are found for some of the matrices $\mathcal{P}_1, \cdots, \mathcal{P}_m$, the equations (6.4) may be particularly useful, not only permitting an easy determination of the remaining z_{ui}, but giving them in the order required by Frobenius' theorem and used in Theorem 6.1.

The matrix $Z = (z_{ui})$ seems deserving of further study. As an indication of its usefulness we make the following remark:

$$\sum_{u=0}^{m} \alpha_u z_{ui} = \operatorname{Tr} B_i = v, \qquad i = 0,$$

$$= 0, \qquad i = 1, 2, \cdots, m.$$

This is equivalent to the system of equations

$$(\alpha_0, \alpha_1, \cdots, \alpha_m)Z = (v, 0, 0, \cdots, 0)$$

providing an alternative to (6.1) for determining α_u.

The authors are thankful to W. S. Connor and Karl Goldberg for several stimulating discussions during the preparation of this paper.

REFERENCES

[1] R. C. Bose, "Least square aspects of analysis of variance," University of North Carolina, Institute of Statistics Mimeograph Series No. 9.

[2] R. C. Bose, "Versuche in unvollständiger Blöcken," Gastvarlesung Universität Frankfurt/M., Naturwissenschaftliche Fakultät, 1955.

[3] R. C. Bose and W. H. Clatworthy, "Some classes of partially balanced designs," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 212–232.

[4] R. C. Bose, W. H. Clatworthy and S. S. Shrikhande, "Tables of partially balanced incomplete block designs with two associate classes," North Carolina Agricultural Experimental Station Technical Bulletin No. 107, 1954.

[5] R. C. Bose and K. R. Nair, "Partially balanced incomplete block designs," *Sankhyā*, Vol. 4 (1939), pp. 337–372.

[6] R. C. Bose and T. Shimamoto, "Classification and analysis of partially balanced incomplete block designs with two associate classes," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–184.

[7] Alfred Brauer, "Limits for the characteristic roots of a matrix. IV: Applications to stochastic matrices," *Duke Math. J.*, Vol. 19 (1952), pp. 75–91.

[8] W. H. Clatworthy, "Contributions on partially balanced incomplete block designs with two associate classes," National Bureau of Standards Applied Math. Series No. 47, 1956.

[9] W. S. Connor and W. H. Clatworthy, "Some theorems for partially balanced designs," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 100–112.

[10] E. C. Dade and K. Goldberg, "Incidence algebras," to be published.

[11] D. Konig, *Theorie der Endlichen und Unendlichen Graphen*, Chelsea, 1950.

[12] R. G. Laha and J. Roy, "Two associate partially balanced designs involving three replications," *Sankhyā*, Vol. 17 (1956), pp. 175–184.

[13] C. C. MacDuffee, *An Introduction to Abstract Algebra*, Wiley, 1940.

[14] C. C. MacDuffee, *The Theory of Matrices*, Chelsea, 1946.

[15] D. M. Mesner, "An investigation of certain combinatorial properties of partially balanced incomplete block experimental designs and association schemes, with a detailed study of designs of Latin square and related types," unpublished doctoral thesis, Michigan State University, 1956.

[16] I. C. Ross and F. Harary, "On the determination of redundancies in sociometric chains," *Psychometrica*, Vol. 17 (1952), pp. 195–208.

[17] W. A. Thompson, Jr., "On the ratio of variances in the mixed incomplete block model," unpublished doctoral thesis, University of North Carolina, 1954.

[18] M. N. Vartak, "On an application of Kronecker product of matrices to statistical designs," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 420–438.

[19] W. A. Thompson, Jr., "A note on P. B. I. B. design matrices," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 919–922.

ON A CHARACTERIZATION OF THE TRIANGULAR ASSOCIATION SCHEME[1]

By S. S. Shrikhande

University of North Carolina

1. Introduction. A partially balanced incomplete block design with two associate classes [1] is said to be triangular [2] if the number of treatments is $v = n(n-1)/2$ and the association scheme is an array of n rows and n columns with the following properties:

(a) The positions in the principal diagonal are blank.

(b) The $n(n-1)/2$ positions above the principal diagonal are filled by the numbers $1, 2, \cdots, n(n-1)/2$ corresponding to the treatments.

(c) The array is symmetric about the principal diagonal.

(d) For any treatment x the first associates are exactly those treatments which lie in the same row and same column as x.

It is then obvious that in the notation of [1]

(1) the number of first associates of any treatment is $n_1 = 2n - 4$;

(2) with respect to any two treatments x_1 and x_2 which are first associates (denoted by $(x_1, x_2) = 1$), the number of treatments which are first associates of both x_1 and x_2 is

$$p_{11}^1(x_1, x_2) = n - 2;$$

(3) with respect to any two treatments x_3 and x_4 which are second associates (denoted by $(x_3, x_4) = 2$) the number of treatments which are first associates of both x_3, x_4 is

$$p_{11}^2(x_3, x_4) = 4.$$

In an interesting paper Connor [3] has shown that if $n \geqq 9$, (1), (2) and (3) above imply (a), (b), (c) and (d), i.e., the association scheme is triangular. In this paper we derive a theorem and utilize it to prove that Connor's result is true for the cases $n = 5, 6$.

2. A characterization of the triangular association scheme. We prove a theorem which is equivalent to the theorem proved by Connor [3].

Theorem: *A necessary and sufficient condition that a partially balanced incomplete block design with two associate classes for $n(n-1)/2$ treatments with $p_{11}^1(x_1, x_2) = n - 2$, where $(x_1, x_2) = 1$, has a triangular association scheme, is that all the first associates of any treatment x whatsoever can be divided into two*

Received July 19, 1958.

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 18(600)-83. Reproduction in whole or in part is permitted for any purpose of the United States Government.

sets $(y_1, y_2, \cdots, y_{n-2})$, *and* $(z_1, z_2, \cdots, z_{n-2})$ *such that* $(y_i, y_j) = (z_i, z_j) = 1$ for $i \neq j = 1, 2, \cdots, n-2$.

PROOF. Necessity is obvious. We now prove the sufficiency.

Since y_i has $(n-3)$ first associates y_j and $p_{11}^1(x, y_i) = n-2$, y_i has just one treatment from the other set, say z_i such that $(y_i, z_i) = 1$ and $(y_i, z_j) = 2$ for $j \neq i$. Now suppose that $(y_{i_1}, z_i) = 1$ for $i_1 \neq i$. Then z_i has y_i, y_{i_1} and z_j, $j \neq i$ for its first associates giving the value $p_{11}^1(x, z_i) = n-1$ which is a contradiction. Hence we can pair off the treatments of the two sets such that

$$(y_i, z_i) = 1, \qquad (y_i, z_j) = 2, \qquad i \neq j = 1, 2, \cdots, n-2.$$

We will use this fact repeatedly. Further it is obvious that if the first associates of any treatment can be divided into sets as above, this division into two sets can be done in a unique manner.

For simplicity let us assume that the first associates of 1 are given by the two sets,

$$\begin{array}{llll} (2, & 3, & \cdots, & (n-1)) \quad \text{and} \\ (n, & n+1, & \cdots, & (2n-3)), \end{array}$$

where any two treatments in the same column are first associates and two treatments from different columns are second associates. We will adopt this method of writing to indicate the relationship of the two treatments from different sets.

We now write the rows,

$$\begin{array}{llllll} * & 1 & 2 & 3 & \cdots & (n-1) \\ 1 & * & n & n+1 & \cdots & (2n-3), \end{array}$$

where $*$ denotes that the corresponding position is blank.

Now amongst the treatments occurring so far the first associates of 2 are $1, 3, 4, \cdots, n-1$ and n. Let the remaining first associates be $(2n-2)$, $(2n-1)$, $\cdots$, $(3n-6)$. Assume without loss of generality that

$$(3, 2n-2) = (4, 2n-1) = \cdots = (n-1, 3n-6) = 1;$$

then we form the third row by putting 2, n and $*$ in the first three positions respectively and placing $(2n-2)$, $(2n-1)$, $\cdots$, $(3n-6)$ below 3, 4, $\cdots$, $(n-1)$ respectively. Thus, we have the three following rows:

$$\begin{array}{llllllll} * & 1 & 2 & 3 & 4 & \cdots & (n-1) \\ 1 & * & n & n+1 & n+2 & \cdots & (2n-3) \\ 2 & n & * & 2n-2 & 2n-1 & \cdots & (3n-6). \end{array}$$

We note that $*$ occurs in the principal diagonal positions and the array written so far is symmetric and that the new first associates of 2 are written after the position of the $*$ in the third row.

Now consider the first associates of treatment 3. The only treatments till now which are first associates of 3 are $1, 2, 4, \cdots, n-1, n+1, 2n-2$. Let the remaining $(n-4)$ first associates be $(3n-5)$, $(3n-4)$, $\cdots$, $(4n-10)$. The

two sets of first associates of 3 are

$$(1, \quad 2, \quad 4, \quad 5, \quad \cdots \quad (n-1))$$

and

$$(n+1, \quad 2n-2, \quad 3n-5, \quad 3n-4, \quad \cdots \quad (4n-10)),$$

where we can assume without loss of generality that the treatments in the same column are first associates. We now write down the fourth row to give

*	1	2	3	4	$\cdots$	$(n-1)$
1	*	n	$n+1$	$n+2$	$\cdots$	$(2n-3)$
2	n	*	$2n-2$	$2n-1$	$\cdots$	$(3n-6)$
3	$n+1$	$2n-2$	*	$3n-5$	$\cdots$	$(4n-10)$.

The same method can be used to write down the other arrays corresponding to $4, 5, \cdots, (n-1)$ respectively. It is easy to see that all the positions above the principal diagonal are filled in with the numbers $1, 2, \cdots, n(n-1)/2$ occurring just once. Thus, conditions (a), (b) and (c) are satisfied. Further, any treatment x occurs just in one position above the principal diagonal, say in row i and column $j(\neq i)$. Then, it also occurs in row j and column i. Hence, the first associates of x are all the treatments of row i and all the treatments of row j. By symmetry the treatments of column j are exactly those occurring in row j. Hence, the first associates of x are exactly those treatments which occur in the same row and same column as x. Thus (d) is also satisfied. Hence the association scheme is triangular. This completes the proof.

As stated previously, this theorem is equivalent to one given by Connor [3]. In the present form, however, it is more directly useful.

3. Uniqueness of the triangular scheme for $n = 5$. LEMMA 1. *The first associates of any treatment whatsoever for the design with parameters,*

$$v = 10, \quad n_1 = 6, \quad n_2 = 3, \tag{3.1}$$

$$p_{ij}^1 = \begin{pmatrix} 3 & 2 \\ & 1 \end{pmatrix}, \qquad p_{ij}^2 = \begin{pmatrix} 4 & 2 \\ & 0 \end{pmatrix},$$

can be divided into two sets of three each such that any two treatments of the same set are first associates.

PROOF: Assume that first associates of 1 are 2, 3, 4, 5, 6, and 7 of which 3, 4 and 5 are first associates of 2, and 6 and 7 are second associates of 2. We then have

$$\begin{aligned} (1,2) = (1,3) = (1,4) = (1,5) = (1,6) = (1,7) &= 1 \\ (2,3) = (2,4) = (2,5) &= 1 \\ (2,6) = (2,7) &= 2. \end{aligned}$$

We show that $(6, 7) = 1$. Suppose not, then $(6, 7) = 2$ and 2 is second associate of both 6 and 7 contradicting $p_{22}^2(6, 7) = 0$. Thus, we must have $(6, 7) = 1$.

Consider the pair $(1, 6) = 1$. 7 is first associate and 2 is second associate of 6. Hence 6 has two first associates and one second associate from the set (3, 4, 5). Assume that $(6, 4) = 2$ and hence, $(6, 3) = (6, 5) = 1$. Similarly, 7 has two first associates and one second associate from the set (3, 4, 5). We show that 4 cannot be second associate of 7. For if $(7, 4) = 2$, then $(6, 7) = 1$ and 2 and 4 are common second associates of both 6 and 7, contradicting $p_{22}^1(6, 7) = 1$. Hence, we must have $(7, 4) = 1$, and hence 7 has one first associate and one second associate from the set (3, 5). We can assume without loss of generality that $(7, 3) = 2$, $(7, 5) = 1$. Then we have the set (5, 6, 7) such that any two treatments of the set are first associates. Now we consider the set (2, 3, 4). We already know that $(2, 3) = (2, 4) = 1$. We now show that $(3, 4) = 1$. Now the common first associates of 1 and 5 are 2, 6, 7. Hence, 3 and 4 are second associates of 5. Hence, we must have $(3, 4) = 1$ in the same way as we obtained $(6, 7) = 1$ above. Thus, (2, 3, 4) is another set of first associates of 1 such that any two members of the set are first associates. A similar result is true for any other treatment. This proves the Lemma.

An appeal to the theorem now gives the corollary:

COROLLARY. *A partially balanced design with parameters* (3.1) *has triangular association scheme.*

4. Uniqueness of the triangular association scheme for $n = 6$. LEMMA 2. *The first associates of any treatment whatsoever for the design with parameters,*

$$v = 15, \qquad n_1 = 8, \qquad n_2 = 6 \tag{4.1}$$

$$p_{ij}^1 = \begin{pmatrix} 4 & 3 \\ 3 & \end{pmatrix}, \qquad p_{ij}^2 = \begin{pmatrix} 4 & 4 \\ 1 & \end{pmatrix},$$

can be broken up into two sets of four each such that any two treatments of the same set are first associates.

PROOF: Assume without loss of generality that the second associates of treatment 1 are the treatments 10, 11, 12, 13, 14, 15, and those of 10 are 1, 4, 5, 8, 9, 15, so that 15 is the only common second associate (since $p_{22}^2 = 1$) of both 1 and 10. Hence, any two treatments of the set (1, 10, 15) are second associates. Then, considering the pairs (1, 10) and (1, 15), it is easy to see that 11, 12, 13, 14 are first associates of both 10 and 15. Now, $(1, 11) = 2$ and 10 and 15 are first associates of 11. Hence, from the value $p_{11}^2(1, 11) = 4$, we see that 11 has two first associates from the set (12, 13, 14). Let these be 12 and 13 so that

$$(11, 12) = (11, 13) = 1, \qquad (11, 14) = 2.$$

Now, $(1, 14) = 2$ and, as before, 14 has two first associates from the set (11, 12, 13). These are obviously 12 and 13, since $(14, 11) = 2$. Hence, we have

$$(12, 14) = (13, 14) = 1.$$

Similarly considering the pair (1, 12) and noting that $(12, 11) = (12, 14) = 1$, we get $(12, 13) = 2$

All the above information can be easily read by writing the second associates of 1 in the following scheme:

$$S_1: \begin{array}{c|ccc} & 1 & 10 & 15 \\ \hline & 11 & & \\ & 14 & & \\ \hline & 12 & & \\ & 13 & & \end{array}.$$

The explanation of the scheme is as follows: Treatments 10, 11, $\cdots$, 15 are second associates of 1, where 11, 12, 13, 14 are first associates and 15 is a second associate of 10. We write 10 and 15 in the row in the second and third positions respectively. Treatments 11, 12, 13, 14 are also first associates of 15. Further any two treatments of the set (1, 10, 15) are second associates. Of the six pairs from the set (11, 12, 13, 14), only those marked by straight lines on the left are second associates, while the remaining four pairs are first associates. The relations implied by S_1 are written completely as follows:

$$\begin{aligned} (4.2)\quad & (1, 10) = (1, 11) = (1, 12) = (1, 13) = (1, 14) = (1, 15) = 2 \\ & (10, 15) = 2, \quad (10, 11) = (10, 12) = (10, 13) = (10, 14) = 1 \\ & (15, 11) = (15, 12) = (15, 13) = (15, 14) = 1 \\ & (11, 14) = 2, \quad (11, 12) = (11, 13) = 1 \\ & (14, 12) = (14, 13) = 1, \quad (12, 13) = 2. \end{aligned}$$

Now, among the seven treatments above the only second associates of 15 are 1 and 10. Let the remaining four second associates of 15 be 2, 3, 6, 7. Then, as before 2, 3, 6, 7 are first associates of both 1 and 10. Without loss of generality assume that $(2, 7) = (3, 6) = 2$ and hence, $(2, 3) = (2, 6) = (7, 3) = (7, 6) = 1$. Hence, we can represent the second associates of 15 in the following scheme:

$$S_2: \begin{array}{c|ccc} & 15 & 1 & 10 \\ \hline & 2 & & \\ & 7 & & \\ \hline & 3 & & \\ & 6 & & \end{array}.$$

The new relations implied by S_2 are

$$\begin{aligned} (4.3)\quad & (15, 2) = (15, 3) = (15, 6) = (15, 7) = 2 \\ & (1, 2) = (1, 3) = (1, 6) = (1, 7) = 1 \\ & (10, 2) = (10, 3) = (10, 6) = (10, 7) = 1 \\ & (2, 7) = 2, \quad (2, 3) = (2, 6) = 1 \\ & (7, 3) = (7, 6) = 1, \quad (3, 6) = 2. \end{aligned}$$

We now consider the relation of any treatment from the set (2, 3, 6, 7) with any treatment of the set (11, 12, 13, 14).

Now, $(1, 2) = 1$ and 10, 15 are, respectively, first and second associates of 2.

Hence, from the value $p_{12}^1(2, 1) = 3$ and $p_{22}^1(2, 1) = 3$, we see that 2 has exactly two first associates and exactly two second associates from the set (11, 12, 13, 14). Suppose we have $(2, 11) = (2, 14) = 1$ and hence, $(2, 12) = (2, 13) = 2$. Then, since $(12, 13) = 2$ and the common second associates of both 12 and 13 are 1 and 2, we get $p_{22}^2 = 2$. Hence, a contradiction. We get a similar contradiction, if we assume that 11, 14 are second associates of 2. Hence, the only possible case is that 2 has just one first associate and just one second associate from each set (11, 14) and (12, 13). We can assume without loss of generality that

$$(2, 11) = (2, 12) = 1, \qquad (2, 14) = (2, 13) = 2. \tag{4.4}$$

Now, consider the pair $(15, 11) = 1$. Here 1 and 10 are respectively second and first associates of 11. Hence, as before, of the remaining four second associates of 15, i.e., 2, 3, 6, 7, exactly two are first associates and exactly two are second associates of 11. A similar argument shows that 11 has exactly one first associate and exactly one second associate from the sets (2, 7) and (3, 6). But, we already have $(11, 2) = 1$, and hence we must have

$$(11, 7) = 2. \tag{4.5}$$

A similar argument considering the pair $(1, 7) = 1$ shows that

$$(7, 14) = 1. \tag{4.6}$$

In the same manner we also get

$$(7, 13) = 1, \qquad (7, 12) = 2. \tag{4.7}$$

We, thus, get the relationship of any treatment from the set (2, 7) with any treatment of the set (11, 12, 13, 14). A similar argument shows that 11 has just one first associate and just one second associate from the set (3, 6). Without loss of generality we can assume that

$$\begin{aligned} &(11, 6) = 1, \qquad (11, 3) = 2, \\ &(14, 3) = 1, \qquad (14, 6) = 2. \end{aligned} \tag{4.8}$$

Now, the relationship of 3, 6 with 12, 13 remains to be determined. Obviously, we have the two following possibilities: Either

$$\text{(A)}: (6, 12) = (3, 13) = 1, \qquad (6, 13) = (3, 12) = 2, \tag{4.9}$$

or

$$\text{(B)}: (6, 12) = (3, 13) = 2, \qquad (6, 13) = (3, 12) = 1. \tag{4.10}$$

We now proceed to show that case (B) is impossible.

Amongst the eleven treatments occurring so far, the only second associates of 10 are 1 and 15. Hence, the remaining second associates of 10 are 4, 5, 8, 9. Of the six possible pairs just two of them are second associates. Assume without loss of generality that

$$\begin{aligned} &(4, 9) = (5, 8) = 2 \\ &(4, 5) = (4, 8) = (9, 5) = (9, 8) = 1. \end{aligned} \tag{4.11}$$

TABLE 1

Treatment	First Associates										Second Associates									
Column 1	Column 2								Column 3	Column 4	Column 5						Column 6		Column 7	
1	2	3	4	5	6	7	8	9			10	11	12	13	14	15				
2	1	3	6	10	11	12					7	13	14	15						
3	1	2	7	10	12	14					6	11	13	15						
4	1	5	8	11	15				13	12	9	10	14				12		13	
5	1	4	9	14	15				13	12	8	10	11				12		13	
6	1	2	7	10	11	13					3	12	14	15						
7	1	3	6	10	13	14					2	11	12	15						
8	1	4	9	11	15				12	13	5	10	14				13		12	
9	1	5	8	14	15				12	13	4	10	11				13		12	
10	2	3	6	7	11	12	13	14			1	4	5	8	9	15				
11	2	4	6	8	10	12	13	15			1	3	5	7	9	14				
12	2	3	10	11	14	15			8	9 4 5	1	6	7	13			4	5	8	9
13	6	7	10	11	14	15			4	5 8 9	1	2	3	12			8	9	4	5
14	3	5	7	9	10	12	13	15			1	2	4	6	8	11				
15	4	5	8	9	11	12	13	14			1	2	3	6	7	10				

We can represent this by

$$S_2: \begin{array}{c} 10 \quad 15 \quad 1 \\ \vert\, 4 \\ \vert\, 9 \\ \vert\, 8 \\ \vert\, 5 \end{array}.$$

Also, we can assume with loss of generality by considering the pair (10, 11) = 1, that

$$(4.12) \qquad \begin{aligned} &(11, 4) = (11, 8) = (14, 9) = (14, 5) = 1 \\ &(11, 9) = (11, 5) = (14, 4) = (14, 8) = 2. \end{aligned}$$

We note that the relations (4.11) and (4.12) do not depend in any manner on the relation (B). We summarize the information given by (4.2), $\cdots$, (4.8), (4.10), (4.11) and (4.12) in Table 1 in columns 1, 2 and 5.

We now consider the possible relationship of the treatments 12, 13 with the treatments 4, 5, 8, 9. We have the four possible cases,

$$\text{(i)} \qquad \begin{aligned} &(12, 4) = (12, 8) = (13, 9) = (13, 5) = 1 \\ &(12, 9) = (12, 5) = (13, 4) = (13, 8) = 2, \end{aligned}$$

$$\text{(ii)} \qquad \begin{aligned} &(12, 9) = (12, 5) = (13, 4) = (13, 8) = 1 \\ &(12, 4) = (12, 8) = (13, 9) = (13, 5) = 2, \end{aligned}$$

$$\text{(iii)} \qquad \begin{aligned} &(12, 8) = (12, 9) = (13, 4) = (13, 5) = 1 \\ &(12, 4) = (12, 5) = (13, 9) = (13, 8) = 2, \end{aligned}$$

$$\text{(iv)} \qquad \begin{aligned} &(12, 8) = (12, 9) = (13, 4) = (13, 5) = 2 \\ &(12, 4) = (12, 5) = (13, 9) = (13, 8) = 1. \end{aligned}$$

Of these case (i) is impossible, since otherwise from Table 1 and columns 1 and 2 we see that $(11, 12) = 1$ and 11 and 12 would have five common first associates contradicting $p_{11}^1 = 4$. Similarly, case (ii) gives $(11, 12) = 1$ and $p_{11}^1(11, 12) = 3$. We are thus left with only case (iii) and case (iv).

We now consider case (iii). The information given by this is entered in columns 3 and 6. We now consider the possible relationships of 2 and 7 with 4, 5, 8, 9. We have the following cases to be considered:

(α) $\quad (2, 9) = (2, 5) = (7, 4) = (7, 8) = 1$
$\quad\quad\quad (2, 4) = (2, 8) = (7, 5) = (7, 9) = 2,$

(β) $\quad (2, 9) = (2, 5) = (7, 4) = (7, 8) = 2$
$\quad\quad\quad (2, 4) = (2, 8) = (7, 5) = (7, 9) = 1,$

(γ) $\quad (2, 9) = (2, 8) = (7, 4) = (7, 5) = 1$
$\quad\quad\quad (2, 4) = (2, 5) = (7, 9) = (7, 8) = 2,$

(δ) $\quad (2, 9) = (2, 8) = (7, 4) = (7, 5) = 2$
$\quad\quad\quad (2, 4) = (2, 5) = (7, 9) = (7, 8) = 1.$

Referring to Table 1 and columns 1, 2, 3, 5 and 6 we see that $(14, 2) = 2$, and cases (α) and (β) give p_{22}^2 $(14, 2) = 2$ and 0 respectively, giving a contradiction since $p_{22}^2 = 1$. Similarly, $(13, 2) = 2$ and (γ) and (δ) give p_{22}^2 $(13, 2) = 0$ and 2 respectively, again a contradiction. Hence, we see that case (iii) is impossible.

We now suppress the information in columns 3 and 6 and put down the information given by case (iv) in columns 4 and 7. With case (iv) we again consider the cases (α), (β), (γ) and (δ). We now look up columns 1, 2, 4, 5, 7 of Table 1. Again, $(14, 2) = 2$ and (α) and (β) give p_{22}^2 $(14, 2) = 2$ and 0 respectively. Similarly, $(13, 2) = 2$ and (γ) and (δ) give p_{22}^2 $(13, 2) = 2$ and 0 respectively. Hence, a contradiction again. Thus, case (iv) is also impossible. It is now clear that case (B) is impossible, and we are left with case (A) alone. The relations (4.2), . . . , (4.9) now give the following two sets of first associates of treatment 10:

$$(11, 12, 2, 6)$$

and

$$(14, 13, 7, 3),$$

where any two treatments from each of the two sets are first associates.

A similar result can be proved for any treatment x by considering its two second associates y and z where $(y, z) = 2$ and taking the four remaining second associates of y and z which will be the eight first associates of x. This completes the proof of Lemma 2.

The application of the theorem now gives the corollary.

COROLLARY. *A design with parameters* (4.1) *has triangular association scheme.*

5. Uniqueness of Triangular Association Scheme for $n \geqq 9$. A lemma similar to Lemmas 1 and 2 can be proved for this case which implies that the association

scheme is triangular if $n \geqq 9$. The proof is omitted, as another proof has already been given by Connor [3].

REFERENCES

[1] R. C. Bose and K. R. Nair, "Partially balanced incomplete block designs", *Sankhya*, Vol. 4 (1939), pp. 337–372.

[2] R. C. Bose and T. Shimamoto, "Classification and analysis of partially balanced designs with two associate classes", *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–190.

[3] W. S. Connor, "The uniqueness of the triangular association scheme", *Ann. Math. Stat.*, Vol. 29 (1958), pp. 262–266.

ON A GENERALISATION OF THE KRONECKER PRODUCT DESIGNS[1]

By B. V. Shah

University of Bombay

1. Summary and introduction. The use of incomplete block designs for estimating and judging the significance of the difference of treatment effects is now standard. Fisher and Yates [2] have provided a complete table of balanced incomplete block designs (BIB) for the value of $r, k \leqq 10$. In this paper a method of constructing a special class of partially balanced incomplete block designs (PBIB) from the known BIB designs is given. Vartak [6] and Sillito [4] have used the Kronecker product of matrices to construct statistical designs. Their method and the method given in this paper differ only in the fact that in their method two distinct elements of a matrix are replaced by two distinct matrices, while in the method considered in this paper, two or more distinct elements of a matrix are replaced by different matrices. All the PBIB designs considered in this paper are with three associate classes, and the rectangular association scheme for $v = pq$ treatments are written in p rows and q columns so that treatments in the same row are the first associates, treatments in the same column are the second associates and the rest are the third associates.

2. Notation and some definitions. Throughout this paper, parameters of a BIB design will be denoted by $v^*, b^*, r^*, k^*, \lambda^*$. The parameters of a PBIB will be denoted by $v, b, r, k, \lambda_1, \cdots, \lambda_m, n_1, \cdots, n_m, P^*_{ij}$, as defined by Bose and Nair [1] and later generalised by Nair and Rao [3].

The following three designs, with the given incidence matrices and parameters will be included in the BIB's in this paper.

(a) A null design with the incidence matrix, $0(v^*, b^*)$ [the $v^* \times b^*$ matrix with all the elements equal to zero]. Parameters: $v^*, b^*, r^* = k^* = \lambda^* = 0$.

(b) A randomised block design with the incidence matrix $E(v^*, b^*)$ [the $v^* \times b^*$ matrix with all the elements equal to unity]. Parameters: $v^*, b^*, r^* = \lambda^* = b^*, k^* = v^*$.

(c) A design with the incidence matrix $I(v^*)$ [the $v^* \times v^*$ identity matrix]. Parameters: $v^* = b^*, r^* = k^* = 1, \lambda^* = 0$.

Definition 2.1. Two BIB designs with incidence matrices N_1 and N_2 and parameters $v_1^*, b_1^*, r_1^*, k_1^*, \lambda_1^*$ and $v_2^*, b_2^*, r_2^*, k_2^*, \lambda_2^*$, respectively will be called associable designs, if and only if $b_1^* = b_2^*, v_1^* = v_2^*$ and the design $N_{12} = [N_1/N_2]$ formed by joining the corresponding blocks of the two designs has the following properties:

(i) The ith treatment of N_1 occurs with the ith treatment of N_2 exactly μ_{12} times for all $i = 1, 2, \cdots, v_1^*$.

Received March 3, 1958; revised August 1, 1958.

[1] This work was supported by a Research Training Scholarship of the Government of India.

(ii) The ith treatment of N_1 occurs with the jth treatment of N_2 exactly η_{12} times for all $i \neq j; i, j = 1, 2, \cdots, v_1^*$.

The following results are direct consequences of the above definition:

A BIB design (N_1) with parameters $v^*, b^*, r^*, k^*, \lambda^*$ is associable with

(i) a null design $\{N_2 = 0(v^*, b^*)\}$ and $\mu_{12} = 0 = \eta_{12}$

(ii) a randomised block design $\{N_3 = E(v^*, b^*)\}$ and $\mu_{13} = r^* = \eta_{13}$,

(iii) its complementary design $\{N_4 = E(v^*, b^*) - N_1\}$ and $\mu_{14} = 0$ $\eta_{14} = r^* - \lambda^*$, and

(iv) itself (N_1) and $\mu_{11} = r^*$ $\eta_{11} = \lambda^*$.

DEFINITION 2.2. Let $A = a_{ij}$, $i = 1, 2, \cdots, m, j = 1, 2, \cdots, n$, be a matrix whose elements take the S values $1, 2, \cdots, S$. We shall call matrix A a balanced matrix in S integers if and only if the following conditions are satisfied:

(i) The number of times the integer p occurs in a column is the same for all the columns and is equal to α_p, say.

(ii) The number of times the integer p occurs in a row is the same for all the rows and is equal to β_p, say.

(iii) The number of times the combination $\binom{p}{q}$ or $\binom{q}{p}$ occurs in a pair of rows is the same for all the pairs of rows and is equal to γ_{pq}, say. ($\binom{p}{q}$ and $\binom{q}{p}$ will be considered as identical combinations).

From Definition 2.2 we can easily prove the following lemmas:

LEMMA 2.1. *If, in a balanced matrix A in S integers, some of the integers, say h, are replaced by 1 and the remaining $(S - h)$ integers by 0, then the matrix A will be converted into an incidence matrix of a BIB design.*

LEMMA 2.2. *If A is a matrix whose elements take the S values $1, 2, \cdots, S$ and if any one of the integers is replaced by 1 and the remaining $S - 1$ by 0, or any two of the integers are replaced by 1 and remaining $S - 2$ are replaced by 0 and if, in all these $S + \binom{S}{2}$ ways, it is found that the matrix A is converted into an incidence matrix of a BIB design, then the matrix A must be a balanced matrix in S integers.*

COROLLARY TO LEMMAS 2.1 AND 2.2. *The necessary and sufficient condition for a matrix A to be balanced is that there exist S BIB designs $N_1, N_2, \cdots, N_S$ such that $\sum N_i = E(m, n)$, $(N_i + N_j)$ is also an incidence matrix of a BIB design for all $i \neq j$ and $A = \sum_{i=1}^{S} iN_i$.*

LEMMA 2.3. (Generalisation of Lemma 2.1). *If there is a balanced matrix A in S_2 integers, and these S_2 integers are divided in S_1 $(S_1 < S_2)$ groups such that each group contains at least one integer and all the elements of a group are replaced by an identical integer, then the matrix A will become a balanced matrix in S_1 integers.*

It is obvious that the incidence matrix of a BIB is a balanced matrix in two integers. Further, from the corollary to Lemmas 2.1 and 2.2 for the existence of a balanced matrix in S integers with parameters as given in Definition 2.2, it is necessary that several BIB's with $v^* = m$, $b^* = n$ and for all the values of $k^* = \alpha_i$, or $\alpha_i + \alpha_j, j, i = 1, 2, \cdots, S$ exist separately. The existing BIB's satis-

fying the above conditions and $v, b \leqq 15$ for $S \geqq 3$ are as follows:

	$v^* = m$	$b^* = n$	α's
(1)	3	3, 6, 9, 12 or 15.	$\alpha_1 = \alpha_2 = \alpha_3 = 1$
(2)	4	12	any values such that $\sum \alpha_i = 4$
(3)	5	10	any values such that $\sum \alpha_i = 5$
(4)	6	15	$\alpha_1 = \alpha_2 = \alpha_3 = 2$
(5)	7	7, 14	$\alpha_1 = \alpha_2 = 3, \alpha_3 = 1$
(6)	9	12	$\alpha_1 = \alpha_2 = \alpha_3 = 3$
(7)	11	11	$\alpha_1 = \alpha_2 = 5, \alpha_3 = 1$
(8)	15	15	$\alpha_1 = \alpha_2 = 7, \alpha_3 = 1$

The balanced matrices with the above parameters are constructible and are given in the appendix. Any other balanced matrix with $m, n \leqq 15$ and $S \geqq 3$ does not exist because the corresponding BIB's do not exist.

3. Construction of PBIB's. Using associable BIB's and a balanced matrix we can construct PBIB designs with three associate classes as given by the following theorem.

THEOREM 3.1. *Let there be S BIB designs with incidence matrices $N_1, N_2, \cdots, N_S$. Let the parameters of the ith design be v^*, b^*, r_i^*, k_i^*, and λ_i^*. Let the pth design be associable with the qth design with the parameters μ_{pq} and η_{pq} for all $p, q = 1, 2, \cdots, S$. Now let there be an $m \times n$ balanced matrix A in S integers $1, 2, \cdots, S$ with parameters $\alpha_p, \beta_p, \gamma_{pq}$ as defined in 2.2. Now if we replace the integer p in matrix A by the matrix N_p $(p = 1, 2, \cdots, S)$, then matrix A will be converted into an incidence matrix of a PBIB with the following parameters:*

$$v = m \cdot v^*,$$

$$b = n \cdot b^*,$$

$$r = \sum_{p=1}^{S} \beta_p \gamma_p^*,$$

$$k = \sum_{p=1}^{S} \alpha_p k_p^*,$$

$$n_1 = v^* - 1, \qquad n_2 = m - 1, \qquad n_3 = n_1 n_2,$$

$$\lambda_1 = \sum_{p=1}^{S} \beta_p \lambda_p^*, \qquad \lambda_2 = \sum_{p \geqq q=1}^{S} \gamma_{ph} \mu_{pq}, \qquad \lambda_3 = \sum_{p \geqq q=1}^{S} \gamma_{pq} \eta_{pq};$$

$$P_{ij}^1 = \begin{bmatrix} n_1 - 1 & 0 & 0 \\ & 0 & n_2 \\ & & (n_1 - 1)n_2 \end{bmatrix}, \qquad P_{ij}^2 = \begin{bmatrix} 0 & 0 & n_1 \\ & n_2 - 1 & 0 \\ & & (n_2 - 1)n_1 \end{bmatrix},$$

$$P_{ij}^3 = \begin{bmatrix} 0 & 1 & n_1 - 1 \\ & 0 & n_2 - 1 \\ & & (n_1 - 1)(n_2 - 1) \end{bmatrix}.$$

PROOF: Expressions for v, b, r, k are quite obvious and need no proof. Others can be proved as follows:

From the method of construction, it is obvious that we have m groups of v^* treatments each corresponding to a column of matrix A. The treatments belonging to the same group will be called the first associates. The ith treatment of one group and the ith treatment of another group will be called the second associates. The ith treatment of one group and the jth treatment ($i \neq j$) of another group will be called the third associates. Expressions for n_i's and P_{ij}^k's immediately follow from this association scheme.

Now in any row of the matrix A the integer p occurs β_p times, therefore N_p also occurs β_p times. In N_p any pair of treatments occurs together λ_p^* times. Hence any pair of treatments belonging to the same group occurs exactly $\sum_{p=1}^{s} \beta_p \lambda_p^*$ times. Therefore

$$\lambda_1 = \sum_{p=1}^{s} \beta_p \lambda_p^*.$$

Similarly, for any pair of rows of the matrix A the combination $\binom{p}{q}$ or $\binom{q}{p}$ occurs exactly γ_{pq} times. Hence, the ith treatment of one group and the ith treatment of another group together $\sum \gamma_{pq}\mu_{pq}$ times, and the ith treatment of one group and the jth treatment ($i \neq j$) of another group occur together exactly $\sum \gamma_{pq}\, \eta_{pq}$ times. Hence

$$\lambda_2 = \sum_{p \geq q=1}^{s} \gamma_{pq}\, \mu_{pq}$$

and

$$\lambda_3 = \sum_{p \geq q=1}^{s} \gamma_{pq}\, \eta_{pq}.$$

This completes the proof of Theorem 3.1

As an illustration let us take the matrix N_1 as the incidence matrix of a BIB with the parameters v^*, b^*, r^*, k^*, λ^*, N_2 as the incidence matrix of its complementary design ($N_2 = E(v^*, b^*) - N_1$), $N_3 = 0(v^*, b^*)$ and

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix}$$

Then

$$\mu_{12} = 0, \quad \mu_{13} = 0, \quad \mu_{23} = 0,$$
$$\eta_{12} = r^* - \lambda^*, \quad \eta_{13} = 0, \quad \eta_{23} = 0.$$

Hence on substituting N_p in place of p in matrix A, ($p = 1, 2, 3$), we obtain the incidence matrix of a PBIB with the parameters

$$v = 3v^*, \quad b = 3b^*, \quad r = b^*, \quad k = v^*,$$
$$\lambda_1 = b^* - 2r^* + 2\lambda^*, \quad \lambda_2 = 0, \quad \lambda_3 = r^* - \lambda^*, \quad \text{etc.}$$

4. Some particular cases. Let a balanced matrix A in the two integers 0, 1, be the incidence matrix of a BIB with the parameters v_1^*, b_1^*, r_1^*, k_1^*, λ_1^*, and let N_1 be the incidence matrix of a BIB with the parameters v_2^*, b_2^*, r_2^*, k_2^*, λ_2^*, then by taking N_0 as $0(v_2^*, b_2^*)$, $E(v_2^*, b_2^*)$ and $E(v_2^*, b_2^*) - N_1$ respectively we have the following three cases:

i) $N_0 = 0(v_2^*, b_2^*)$, we obtain a PBIB with the parameters

$$v = v_1^* v_2^*, \qquad b = b_1^* b_2^*, \qquad r = r_1^* r_2^*, \qquad k = k_1^* k_2^*,$$

$$\lambda_1 = r_1^* \lambda_2^*, \qquad \lambda_2 = r_2^* \lambda_1^*, \qquad \lambda_3 = \lambda_1^* \lambda_2^*.$$

ii) $N_0 = E(v_2^*, b_2^*)$, we obtain a PBIB with the parameters

$$v = v_1^* v_2^*, \qquad b = b_1^* b_2^*, \qquad r = r_2^* r_2^* + (b_1^* - r_1^*) b_2^*,$$

$$k = k_1^* k_2^* + (v_1^* - k_1^*) v_2^*,$$

$$\lambda_1 = r_1^* \lambda_2^* + (b_1^* - r_1^*) b_2^*,$$

$$\lambda_2 = (b_1^* - 2r_1^* + \lambda_1^*) b_2^* + 2(r_1^* - \lambda_1^*) r_2^* + \lambda_1^* r_2^*,$$

$$\lambda_3 = (b_1^* - 2r_1^* + \lambda_1^*) b_2^* + 2(r_1^* - \lambda_1^*) r_2^* + \lambda_1^* \lambda_2^*.$$

iii) $N_0 = E(v_2^*, b_2^*) - N_1$, we obtain a PBIB with the parameters

$$v = v_1^* v_2^*, \qquad b = b_1^* b_2^*,$$

$$r = r_1^* r_2^* + (b_1^* - r_1^*)(b_2^* - r_2^*),$$

$$k = k_1^* k_2^* + (v_1^* - k_1^*)(v_2^* - k_2^*),$$

$$\lambda_1 = r_1^* \lambda_2^* + (b_1^* - r_1^*)(b_2^* - 2r_2^* + \lambda_2^*),$$

$$\lambda_2 = r_2^* \lambda_1^* + (b_2^* - r_2^*)(b_1^* - 2r_1^* + \lambda_1^*),$$

$$\lambda_3 = (b_2^* - 2r_2^* + \lambda_2^*)(b_1^* - 2r_1^* + \lambda_1^*) + 2(r_2^* - \lambda_2^*)(r_1^* - \lambda_1^*) + \lambda_1^* \lambda_2^*.$$

It will be noticed that case (i) above is exactly identical with the Kronecker Product of BIB's as given by M. N. Vartak [6]; the same method also gives some of the designs by D. A. Sprott [5].

The method employed by G. P. Sillito [4] to construct BIB is identical with the case (iii) above with the conditions $A = N_1$ and $b_1^* = 4(r_1^* - \lambda_1^*)$. Also, of the two PBIB's with $v = b = pq$ given by R. C. Bose [1], one is a particular case of (iii) above with $A = I(p)$ and $N_1 = E(q, q) - I(q)$ and the other is identical with (ii) above on taking $A = E(p, p) - I(p)$ and $N_1 = I(q)$.

Furthermore it should be noted that if there are two designs of this type with the same number of treatments, the same association scheme for three associate classes, and the same block size, we can construct a new design of this type, just by adjoining the two designs.

Lastly if in a design $\lambda_1 = \lambda_3$ or $\lambda_2 = \lambda_3$, the design is reduced to a group divisible design, and if $\lambda_1 = \lambda_2 = \lambda_3$, the design is reduced to a BIB.

5. Acknowledgement. The author is grateful to Prof. M. C. Chakrabarti for his help and guidance. The author is indebted to one of the referees for his valuable suggestions; in particular the balanced matrices (4), (6), (7) and (8) of the appendix are his.

APPENDIX

The following is the complete list of the balanced matrices with $m, n \leqq 15$ and $S \geqq 3$.

Balanced matrices with $m = 3$, $n = 6$, 9 or 12 and $m = 7$, $n = 14$ can be obtained by replicating the solutions (1) and (5) Balanced matrices with $m = 4$, $S = 3$ and $m = 5$, $S = 3, 4$ can be obtained from the solutions (2) and (3) by the help of Lemma 2.4.

(1) $m = n = S = 3$.

1 2 3
2 3 1
3 1 2

(2) $m = S = 4$, $n = 12$.

1 2 3 4 1 2 3 4 1 2 3 4
2 1 4 3 3 4 1 2 4 3 2 1
3 4 1 2 4 3 2 1 2 1 4 3
4 3 2 1 2 1 4 3 3 4 2 1

(3) $m = S = 5$, $n = 10$.

1 2 3 4 5 1 2 3 4 5
2 3 4 5 1 4 5 1 2 3
3 4 5 1 2 2 3 4 5 1
4 5 1 2 3 5 1 2 3 4
5 1 2 3 4 3 4 5 1 2

(4) $m = 6$, $n = 15$, $S = 3$.

1 1 1 1 1 2 2 2 2 2 3 3 3 3 3
1 2 2 2 2 2 3 3 3 3 1 1 1 1 3
2 1 2 3 3 3 1 2 1 3 1 2 2 3 1
2 3 3 2 1 3 2 1 3 1 2 1 3 2 1
3 2 3 1 3 1 1 3 2 1 2 3 2 1 2
3 3 1 3 2 1 3 1 1 2 3 2 1 2 2

(5) $m = n = 7$, $S = 3$.

1 3 2 2 1 2 1
First row;
matrix is cyclic.

(6) $m = 9$, $n = 12$, $S = 3$.

1 2 3 1 2 3 1 2 3 1 2 3
1 2 3 2 3 1 2 3 1 3 1 2
1 2 3 3 1 2 3 1 2 2 3 1
2 3 1 1 2 3 2 3 1 2 3 1
2 3 1 2 3 1 3 1 2 1 2 3
2 3 1 3 1 2 1 2 3 3 1 2
3 1 2 1 2 3 3 1 2 3 1 2
3 1 2 2 3 1 1 2 3 2 3 1
3 1 2 3 1 2 2 3 1 1 2 3

(7) $m = n = 11$, $S = 3$.

3 1 2 1 1 1 2 2 2 1 2

First row;
matrix is cyclic

(8) $m = n = 15$, $S = 3$.

3	1 1 1 1 1 1 1	2 2 2 2 2 2 2
2	3 2 2 1 2 1 1	2 1 2 2 1 1 1
2		
2		
2		
2		
2		
1	1 2 2 2 1 1 2	3 1 1 2 1 2 2
1		
1		
1		
1		
1		
1		

Each 7×7 block is cyclic.

REFERENCES

[1] R. C. Bose and K. R. Nair, "Partially balanced designs," *Sankhya* Vol. 4 (1939), pp. 337–372.

[2] R. A. Fisher and F. Yates, *Statistical Tables*, Oliver and Boyd, 1938.

[3] K. R. Nair and C. R. Rao, "A note on partially balanced incomplete block designs," *Science and Culture*, Vol. 7(1941–42), pp. 568–569.

[4] G. P. Sillito, "An extension property of a class of balanced incomplete block designs," *Biometrika*, Vol. 44 (1957), pp. 278–279.

[5] D. A. Sprott, "Some series of partially balanced incomplete block designs," *Canadian Jour. of Math.*, Vol. 7 (1955), pp. 369–381.

[6] M. N. Vartak, "On an application of Kronecker product of matrices to statistical designs," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 420–438.

APPLICATION OF A MEASURE OF INFORMATION TO THE DESIGN AND COMPARISON OF REGRESSION EXPERIMENTS

By M. Stone

British Medical Research Council Applied Psychology Unit

1. Introduction and Summary. A normal regression experiment can be represented by

$$Y_i = \sum_{j=1}^{k} X_{ij}\,\theta_j + \eta_i \qquad (i = 1, \cdots, n) \tag{1.1}$$

where $\{\eta_i / i = 1, \cdots, n\}$ is a set of normally distributed random variables with zero means and non-singular dispersion matrix C, $\theta = (\theta_1, \cdots, \theta_k)$ is the parameter-vector of interest and $X = (X_{ij})$ is a known $n \times k$ matrix which will be called the allocation matrix. The rows of X will be called the allocation vectors. We denote the experiment by $\mathcal{E}(X, C)$. We assume that C is known; generally it will be a function of X, $C(X)$. The particular realisation of Y will be denoted y. The matrix $F = X'C^{-1}X$ is the Fisher-information-matrix of $\mathcal{E}(X, C)$.

When F is non-singular, one answer to the question "What information does y give about θ?" is to quote F^{-1}, the dispersion matrix of the maximum-likelihood-estimates of θ. A strong argument in favour of this is that F^{-1} is independent of both θ and y. The fact that it is independent of θ means that the answer is not "local"; the fact that it is independent of y leads to simplicity. This approach is taken by Box and Hunter [1] in their work on rotatable designs. However, we must accept the fact that many experimenters wish to have a one-dimensional answer to the question i.e. we must associate with $\mathcal{E}(X, C)$ a single number which we call the "information". For instance Elfving [5] has developed the use of trace F^{-1}. In this paper we adopt the measure of information introduced by Lindley [7]. In Section 2 we generalise Lindley's treatment of the regression situation to include the singular case, explain the uses of the measure and compare it with that of Elfving. Section 3 deals with the analogue of Elfving's main theorem. Theorems 4.1 and 4.2 of Section 4 provide links with the traditional variance approach. In Section 5 we derive the asymptotic form of the measure as the n of (1.1) increases and show that this form can be derived also from Neyman-Pearsonian theory. In Section 6 the influence of nuisance parameters is discussed and an analogue of a theorem of Chernoff [2] is established.

2. The information measure is defined in the Bayesian framework. Generally, if before experimentation we express our knowledge of θ by the prior distribution $p(\theta)$ and, after the experiment defined by the set of probability density functions $p(y/\theta)$, we obtain a posterior distribution $p(\theta/y)$, then the gain of information is

Received November 4, 1957; revised March 15, 1958.

defined as the functional

$$\Delta I = Ip(\theta/y) - Ip(\theta) \tag{2.1}$$

where $Ip(u) = \int p(u) \log p(u)\, du$.

LEMMA 2.1. *If, before $\mathcal{E}(X, C)$, θ is normally distributed with mean μ and non-singular dispersion matrix A then*

$$\Delta I = \tfrac{1}{2} \log | I + AF | . \tag{2.2}$$

PROOF.

$$\begin{aligned} Ip(\theta) &= \int p(\theta) \log [(2\pi)^{-\frac{1}{2}k} | A |^{-\frac{1}{2}} \exp \{-\tfrac{1}{2}(\theta - \mu)'A^{-1}(\theta - \mu)\}]\, d\theta \\ &= -\tfrac{1}{2} \log [(2\pi)^k | A |] - \tfrac{1}{2} \int p(\theta)(\theta - \mu)'A^{-1}(\theta - \mu)\, d\theta \\ &= -\tfrac{1}{2} \log [(2\pi)^k | A |] - \tfrac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} A^{ij} A_{ij} \\ &= -\tfrac{1}{2} \log [(2\pi)^k | A |] - \tfrac{1}{2}k. \end{aligned}$$

Also

$$\begin{aligned} p(\theta/y) &= p(y/\theta)p(\theta)/p(y) \\ &\propto \exp [-\tfrac{1}{2}(y - X\theta)'C^{-1}(y - X\theta) - \tfrac{1}{2}(\theta - \mu)'A^{-1}(\theta - \mu)]. \end{aligned}$$

Therefore $p(\theta/y)$ is normal with dispersion matrix

$$(X'C^{-1}X + A^{-1})^{-1} = (F + A^{-1})^{-1}.$$

Hence

$$Ip(\theta/y) = -\tfrac{1}{2} \log [(2\pi)^k / | F + A^{-1} |] - \tfrac{1}{2}k$$

and

$$\Delta I = \tfrac{1}{2} \log (| F + A^{-1} | | A |) = \tfrac{1}{2} \log | I + AF | .$$

Among the class of regression experiments for which it is reasonable to take a normal prior distribution for θ with dispersion matrix A, the expression for ΔI just derived proves useful in three ways:

Use 1. If we decide to experiment until the gain of information reaches a certain level then the fact that ΔI is independent of y allows us to state in advance whether a particular experiment will give us the required gain. Among experiments which do, we may choose the one which is most economical in some sense. This is the case of fixed-sample-size-experimentation.

Use 2. Two experiments can be compared in the following sense:

"Any result of $\mathcal{E}(X_1, C_1)$ will give $\Delta I_1 - \Delta I_2$ more information than any result of $\mathcal{E}(X_2, C_2)$". We may note that we are not obliged to use the average gain of information (as defined by Lindley [7]) to compare $\mathcal{E}(X_1, C_1)$ and $\mathcal{E}(X_2, C_2)$ although the result of doing so would be the same.

Use 3. If we have a choice of performing any experiment from a given class, we may choose the $\mathcal{E}(X, C)$ which maximises ΔI. ΔI possesses two advantages over the measure trace F^{-1}:

(i) If $\phi = M\theta$ is a non-singular linear transformation then ΔI is the same whether we consider information about θ or about ϕ. For

$$\begin{aligned} &\tfrac{1}{2}\log|I + AF| = \tfrac{1}{2}\log[|M||I + AF||M^{-1}|] \\ &\qquad = \tfrac{1}{2}\log|I + (MAM')(XM^{-1})'C^{-1}(XM^{-1})|. \end{aligned} \tag{2.3}$$

But MAM' is the prior dispersion matrix for ϕ and, under the transformation, $\mathcal{E}(X, C) \to \mathcal{E}(XM^{-1}, C)$ so that (2.3) is the ΔI for ϕ.

(ii) ΔI can be used even when not all the θ_i are estimable i.e. when F is singular, whereas in this case trace F^{-1} becomes infinite. For $|I + AF| = |A||A^{-1} + F|$, A^{-1} is positive-definite and, although F is singular, it remains positive-semi-definite. Hence $|A^{-1} + F|$ and therefore $|I + AF|$ are non-zero.

3. In connection with Use 3 we proceed to show that a theorem proved by Elfving using trace F^{-1} still holds if we adopt ΔI. The theorem is concerned with the following problem: "Given g possible allocation vectors $x(1), \cdots, x(g)$ which are linearly independent, we are to make n independent observations where n is large and each observation can be made at any of the allocation vectors. How are the observations to be allocated to maximise ΔI?". To answer this we need a lemma.

LEMMA 3.1. *If $F(n_1, \cdots, n_g)$ is the Fisher-information-matrix of the experiment consisting of n_i observations at $x(i)$ $(i = 1, \cdots, g)$ with the errors (η) uncorrelated and if we replace n_i in this matrix by np_i, $p_i \geqq 0$, $\sum p_i = 1$, to obtain the matrix $F^*[np_1, \cdots, np_g]$ then in general $|I + AF^*|$ is maximum when no more than $\frac{1}{2}k(k+1)$ of the p_i's are non-zero.*

PROOF. We have assumed that the non-diagonal elements of $C(X)$ are zero. There is no further loss of generality in assuming that $C(X) = I$ for we can always write $\lambda_i\, x(i)$ for $x(i)$ so that this is so. We find that the (r, s)th element of $F(n_1, \cdots, n_g)$ is $\sum_{i=1}^{g} n_i\, x_r(i)x_s(i)$. Then

$$F^*_{rs} = \sum_{i=1}^{g} np_i\, x_r(i)x_s(i).$$

There are two possibilities:

(a) If there exists an i such that $|I + AF^*|$ is maximum when $p_i = 1$ then since $1 \leqq \frac{1}{2}k(k+1)$ the lemma holds.

(b) If $|I + AF^*|$ is maximum at $p = p(m)$ when more than one $p_i(m)$ is non-zero, we proceed as follows.

$$\begin{aligned} \frac{\partial}{\partial p_i}|I + AF^*| &= |A|\,\frac{\partial}{\partial p_i}|A^{-1} + F^*| \\ &= |A|\sum \begin{vmatrix} nx_1^2(i) & nx_1(i)x_2(i) & \cdots \\ A^{21} + F^*_{21} & A^{22} + F^*_{22} & \cdots \\ \vdots & \vdots & \end{vmatrix} \end{aligned}$$

for $i = 1, \cdots, g$ after row-by-row differentiation. Expanding each of these determinants about the row which has been differentiated we have

$$\frac{\partial}{\partial p_i} | I + AF^* | = n | A | x'(i)Q[p]x(i) \qquad (i = 1, \cdots, g)$$

where $Q[p] = \text{adj } (A^{-1} + F^*)$. Now $\sum p_i = 1$ so that, applying the method of undetermined multipliers to variations of the p_i's for i's for which $p_i(m) \neq 0$, we see that there exists a λ such that

$$\left\{ \frac{\partial}{\partial p_i} | I + AF^* | \right\}_{p(m)} = \lambda$$

for i such that $p_i(m) \neq 0$. Hence for such i

$$x'(i)\, Q[p(m)]x(i) = \lambda' \quad \text{where} \quad \lambda' = \lambda/n | A | . \tag{3.1}$$

Now $x'Q[p(m)]x = \lambda'$ is the equation of a central quadric. In general not more than $\frac{1}{2}k(k+1)$ of the given allocation vectors can lie on a central quadric and hence the lemma is established. To make the lemma fully rigorous, we would need to consider the possibility that more than $\frac{1}{2}k(k+1)$ of the vectors do lie on a central quadric. We omit this consideration since it is rather tedious.

The lemma relates directly to the problem stated. From (2.2)

$$\Delta I = \tfrac{1}{2} \log |I + AF(n_1, \cdots, n_g)| = \tfrac{1}{2} \log |I + AF^*[np_1, \cdots, np_g]|$$

evaluated at $p_i = n_i/n$. So that if we do the experiment consisting of $[np_i(m)]$ observations at $x(i)$ $(i = 1, \cdots, g)$, which involves using at most $\frac{1}{2}k(k+1)$ of $x(1), \cdots, x(g)$, for n large we will be making $n - f$ observations in all where $f \leqq g$. The allocation indicated will provide (asymptotically) the maximum ΔI. Thus:

THEOREM 3.1. *To achieve maximum ΔI for the above problem it is not necessary to use more than $\frac{1}{2}k(k+1)$ of the given allocation vectors.*

For n not large the theorem is not necessarily true. Although it does not specify $p(m)$, it is nevertheless useful in providing a rule for rejecting some experiments for using too many allocation vectors. Generally the calculation of $p(m)$ is not feasible. However when $k = 2$ and the elements of $(AF^*[np(m)])^{-1}$ are small, we may proceed to obtain $p(m)$ approximately. By Theorem 3.1 only three allocation vectors need be used. Consider them as the only vectors given: $x(1)$, $x(2)$, $x(3)$. Approximately

$$Q[p] = \begin{pmatrix} \sum np_j x_2^2(j) & -\sum np_j x_1(j)x_2(j) \\ -\sum np_j x_2(j)x_1(j) & \sum np_j x_1^2(j) \end{pmatrix}$$

and hence

$$x'(i)Q[p]x(i)/n = x_1^2(i)\left[\sum p_j x_2^2(j)\right] - 2x_1(i)x_2(i)\left[\sum p_j x_1(j)x_2(j)\right] + x_2^2(i)\left[\sum p_j x_1^2(j)\right] = \sum_{j=1}^{3} p_j E_{ij} \quad (i = 1, 2, 3)$$

where $E_{ij} = [x_1(i)x_2(j) - x_2(i)x_1(j)]^2$. We note that $E_{ij} > 0$ when $i \neq j$ since the vectors $x(i)$ are supposed independent. Then, if $p_i(m) \neq 0$ $(i = 1, 2, 3)$, equation (3.1) gives

$$\begin{pmatrix} 0 & E_{12} & E_{13} \\ E_{21} & 0 & E_{23} \\ E_{31} & E_{32} & 0 \end{pmatrix} \begin{pmatrix} p_1(m) \\ p_2(m) \\ p_3(m) \end{pmatrix} = \begin{pmatrix} \lambda'' \\ \lambda'' \\ \lambda'' \end{pmatrix}$$

where $\lambda'' = \lambda'/n = \lambda/n^2|A|$. Using $\sum p_i(m) = 1$ we get

$$p_i(m) = E_{jk}(E_{ij} + E_{ik} - E_{jk}) \Big/ \sum_{i=1}^{3} E_{jk}(E_{ij} + E_{ik} - E_{jk}) \tag{3.2}$$

for $i = 1, 2, 3$; $(j, k) = (1, 2, 3) - (i)$. Also

$$\lambda'' = 2E_{23}E_{13}E_{12} \Big/ \sum_{i=1}^{3} E_{jk}(E_{ij} + E_{ik} - E_{jk}). \tag{3.3}$$

Hence for $p_i(m) > 0$ $(i = 1, 2, 3)$ either

$$E_{ij} + E_{ik} > E_{jk} \qquad (i = 1, 2, 3) \tag{3.4}$$

or

$$E_{ij} + E_{ik} < E_{jk} \qquad (i = 1, 2, 3). \tag{3.5}$$

The possibility of (3.5) can be readily excluded. We see that (3.3) and (3.5) imply $\lambda'' < 0$ but since Q is positive-definite $0 < x(i)'Qx(i) = \lambda' = n\lambda''$ or $\lambda'' > 0$, which is a contradiction. Therefore only (3.4) is consistent with $p_i(m) > 0$ $(i = 1, 2, 3)$. Equation (3.4) is not always satisfied by three given vectors. For example if $x(2)$ lies between $x(1)$ and $x(3)$ (in their two-dimensional representation) and

$$|x(2)| < \min(|x(1)|, |x(3)|), \qquad E_{ij} = |x(i) \wedge x(j)|^2 = 4A_{ij}^2$$

where A_{ij} = "area between $x(i)$ and $x(j)$". Clearly $A_{13} > A_{12} + A_{23}$; therefore $A_{13}^2 > A_{12}^2 + A_{23}^2$; therefore $E_{13} > E_{12} + E_{23}$.

However if (3.4) is satisfied the $p(m)$ given by (3.2) is that which approximately maximises ΔI. Also since Q is positive-definite $x'Qx = \lambda'$ is an ellipse, so that we need consider only triples of allocation vectors which lie on central ellipses.

We now evaluate the aymptotic maximum of ΔI:

(a) When (3.4) holds

$$\max \Delta I = \tfrac{1}{2} \log |I + AF^*[np(m)]| \cong \tfrac{1}{2} \log |A| + \tfrac{1}{2} \log |F^*[np(m)]|.$$

Now $|F^*[np]|$ is a homogeneous polynomial of degree two in p_i $(i = 1, 2, 3)$. Therefore

$$|F^*[np]| = \tfrac{1}{2} \sum p_i \frac{\partial}{\partial p_i} |F^*[np]|.$$

But, for the $p(m)$ of (3.2),

$$\left(\frac{\partial}{\partial p_i} \mid F^*[np] \mid\right)_{p(m)} = \frac{\lambda}{\mid A \mid}$$

so that

$$\mid F^*[np(m)] \mid = \tfrac{1}{2}\lambda/\mid A \mid = n^2 E_{23} E_{13} E_{12} / \sum E_{jk}(E_{ij} + E_{ik} - E_{jk}).$$

Hence

$$\max \Delta I \cong \log n + \tfrac{1}{2} \log \mid A \mid + \tfrac{1}{2} \log [E_{23} E_{13} E_{12} / \sum E_{jk}(E_{ij} + E_{ik} - E_{jk})]. \tag{3.6}$$

(b) When (3.4) is not satisfied, one of the vectors must have its associated p_i zero. Suppose $p_3(m) = 0$. Then the equations (3.1) lead to $p_2(m)E_{12} = \lambda'/n$ and $p_1(m)E_{12} = \lambda'/n$ so that $p_1(m) = p_2(m) = \frac{1}{2}$ and $\lambda'/n = \frac{1}{2} E_{12}$ or $\lambda = \frac{1}{2} n^2 \mid A \mid E_{12}$. Hence

$$\begin{aligned} \max \Delta I &\cong \tfrac{1}{2} \log |A| + \tfrac{1}{2} \log |F^*| = \tfrac{1}{2} \log |A| + \tfrac{1}{2} \log (\tfrac{1}{2} \lambda/|A|) \\ &= \log n + \tfrac{1}{2} \log |A| + \tfrac{1}{2} \log (E_{12}/4). \end{aligned} \tag{3.7}$$

In conclusion for the case $k = 2$ we state the experimental rule as follows: "Given g allocation vectors select those triples which obey (3.4) and calculate

$$E_{23} E_{13} E_{12} / \sum E_{jk}(E_{ij} + E_{ik} - E_{jk}).$$

Also select those pairs which are not members of triples obeying (3.4) and calculate $E_{12}/4$. Make the observations at the pair or triple which gives the greatest number, with $n/2$ at each vector for a pair and $np_i(m)$ at $x(i)$ $(i = 1, 2, 3)$ for a triple where $p(m)$ is given by (3.2)."

EXAMPLE 1. $k = 2$; $x(1) = (1, 1)$; $x(2) = (0, 1)$; $x(3) = (1, 0)$. Here $E_{12} = E_{13} = E_{23}$; therefore (3.4) is satisfied and, by (3.2), $p_i(m) = 1/3$ $(i = 1, 2, 3)$.

EXAMPLE 2. k general. Suppose the given allocation vectors lie on the line $x = (1, x, \cdots, x^{k-1})$. This is the case of polynomial regression. $x'Qx = \lambda'$ is a polynomial of degree $2k - 2$ in x. Therefore in general at most $2k - 2$ of the vectors lie on a central quadric and therefore need be used.

In his discussion Elfving considers in detail the case $k = 2$. His solution, i.e. the "best allocation", is rather complicated when three vectors are used. For just two vectors he obtains

$$p_1(m) = |x(2)| \,/\, (|x(1)| + |x(2)|)$$

and

$$p_2(m) = |x(1)| \,/\, (|x(1)| + |x(2)|)$$

which clearly conflicts with $p_1(m) = p_2(m) = \frac{1}{2}$ using ΔI. The reason for the difference becomes clear in the case $x(1) = (c_1, 0)$, $x(2) = (0, c_2)$ for which

$$F^*[np] = \begin{pmatrix} np_1 c_1^2 & 0 \\ 0 & np_2 c_2^2 \end{pmatrix}.$$

The different answers arise because, effectively, we minimise the product of the variances of the maximum-likelihood-estimates of θ_1 and θ_2 while Elfving minimises their sum.

4. In this section we consider pairs of experiments $\mathcal{E}(X_1, C_1)$, $\mathcal{E}(X_2, C_2)$ and prove some theorems relating to the cases when $\mathcal{E}_1$ is always to be preferred to $\mathcal{E}_2$. Write $\Delta I_i(A) = \frac{1}{2} \log |I + AF_i|$ $(i = 1, 2)$.

THEOREM 4.1. *A necessary and sufficient condition that* $\Delta I_1(A) \geqq \Delta I_2(A)$ *for all positive-definite* A *is that* $F_1 - F_2$ *be positive-semi-definite.*

PROOF. Sufficiency. We use the fact that if L and M are positive-definite and $L - M$ is positive-semi-definite then $|L| \geqq |M|$. (Proved by diagonalising L and M.) Put $L = A^{-1} + F_1$ and $M = A^{-1} + F_2$; then if $F_1 - F_2$ is positive-semi-definite $|A^{-1} + F_1| \geqq |A^{-1} + F_2|$ which gives $\Delta I_1(A) \geqq \Delta I_2(A)$.

Necessity. $\Delta I_1(A) \geqq \Delta I_2(A)$ for all positive-definite A implies that $|A^{-1} + F_1| \geqq |A^{-1} + F_2|$ for all positive-definite A^{-1}. Now F_1 and F_2 are positive-semi-definite so that there exists a non-singular P such that $P'F_1P$ and $P'F_2P$ are diagonal.

$$P'F_1P = \begin{pmatrix} d_1(i) & & \\ & \ddots & \\ & & d_k(i) \end{pmatrix} \qquad (i = 1, 2) \tag{4.1}$$

Therefore

$$\left| P'A^{-1}P + \begin{pmatrix} d_1(1) & & \\ & \ddots & \\ & & d_k(1) \end{pmatrix} \right| \geqq \left| P'A^{-1}P + \begin{pmatrix} d_1(2) & & \\ & \ddots & \\ & & d_k(2) \end{pmatrix} \right|$$

where A^{-1}, and hence $P'A^{-1}P$, is arbitrary positive-definite. Taking $P'A^{-1}P$ diagonal with all diagonal elements large except the r'th we deduce $d_r(1) \geqq d_r(2)$. Therefore from equations (4.1)

$$P'(F_1 - F_2)P = \begin{pmatrix} d_1(1) - d_1(2) & & \\ & \ddots & \\ & & d_k(1) - d_k(2) \end{pmatrix}$$

is positive-semi-definite. Therefore $F_1 - F_2$ is positive-semi-definite.

We now give simpler proofs of theorems due to Ehrenfeld [4].

THEOREM 4.2. *If* $v_i(t)$ *is the variance of the maximum-likelihood estimate of* $t'\theta$ *from* $\mathcal{E}_i$ *and if* $F_1 - F_2$ *is positive-semi-definite then* $v_1(t) \leqq v_2(t)$ *for all* t *for which* $t'\theta$ *is estimable from both* $\mathcal{E}_1$ *and* $\mathcal{E}_2$.

PROOF. We know that $v_1 = \eta_1'F_1\eta_1$ and $v_2 = \eta_2'F_2\eta_2$ where η_1 and η_2 are any solutions of

$$F_1\eta_1 = t, \qquad F_2\eta_2 = t. \tag{4.2}$$

We have $\eta_1'(F_1 - F_2)\eta_1 \geqq 0$ or

$$v_1 - \eta_1'F_2\eta_1 \geqq 0 \tag{4.3}$$

while

$$(\eta_1 - \eta_2)'F_2(\eta_1 - \eta_2) = \eta_1'F_2\eta_1 - 2\eta_2'F_2\eta_1 + \eta_2'F_2\eta_2$$
$$\geqq 0.$$

From (4.2) $\eta_2'F_2\eta_1 = \eta_1'F_1\eta_1 = v_1$ therefore

$$\eta_1'F_2\eta_1 - 2v_1 + v_2 \geqq 0. \tag{4.4}$$

Adding (4.3) and (4.4) we obtain $v_1 \leqq v_2$.

THEOREM 4.3. *Given g allocation vectors $x(1), \cdots, x(g)$ and their convex hull $\mathcal{C} = \{\sum \lambda_i x(i) / \sum \lambda_i = 1, \lambda_i \geqq 0\}$, if $F(x_1, \cdots, x_n)$ is the Fisher-information-matrix of the experiment consisting of n independent observations at $x_1, \cdots, x_n$ (with unit error variance) where $x_i \varepsilon \mathcal{C}$ then we may take less than $n + g + 1$ observations at the vertices of $\mathcal{C}$ so that their Fisher-information-matrix, F_V, is such that $F_V - F(x_1, \cdots, x_n)$ is positive-semi-definite.*

PROOF. Suppose $x_i = \sum_{j=1}^{g} \lambda_{ij}x(j)$. For each i we have

$$\sum_{j=1}^{g} \lambda_{ij}x(j)x'(j) - x_ix_i' = \sum_{j=1}^{g} \lambda_{ij}x(j)x'(j) - \sum_{j=1}^{g}\sum_{k=1}^{g} \lambda_{ij}\lambda_{ik}x(j)x'(k)$$

and for any t

$$t'\left[\sum_{j=1}^{g} \lambda_{ij}x(j)x'(j) - x_ix_i'\right]t = \sum_{j=1}^{g} \lambda_{ij}t'x(j)x'(j)t - t'x_ix_i't$$
$$= \sum_{j=1}^{g} \lambda_{ij}\left[t'x(j) - \sum_{j=1}^{g} \lambda_{ij}t'x(j)\right]^2$$
$$\geqq 0.$$

Hence $\sum_{j=1}^{g} \lambda_{ij}x(j)x'(j) - x_ix_i'$ is positive-semi-definite for $i = 1, \cdots, n$. Therefore

$$\sum_{i=1}^{n}\sum_{j=1}^{g} \lambda_{ij}x(j)x'(j) - \sum_{i=1}^{n} x_ix_i'$$

is positive-semi-definite. Therefore

$$\sum_{j=1}^{g}\left(\left[\sum_{i=1}^{n} \lambda_{ij}\right] + 1\right)x(j)x'(j) - F(x_1, \cdots, x_n)$$

is positive-semi-definite where $[a]$ is the integral part of a. We can now identify $F_V = \sum_{j=1}^{g}\left(\left[\sum_{i=1}^{n} \lambda_{ij}\right] + 1\right)x(j)x'(j)$ for it is the Fisher-information-matrix of the experiment in which $\left[\sum_{i=1}^{n} \lambda_{ij}\right] + 1$ independent observations are made at $x(j)$ $(j = 1, \cdots, g)$ with error variances unity and also

$$\sum_{j=1}^{g}\left(\left[\sum_{i=1}^{n} \lambda_{ij}\right] + 1\right) \leqq n + g.$$

5. From Section 4 we see that the only case in which the ordering of two experiments by the criterion $\Delta I(A)$ is the same for all A is when $F_1 - F_2$ is

positive-semi-definite. Clearly since $F_1 - F_2$ may be neither positive- nor negative-semi-definite, not all pairs of experiments can be compared in this clear-cut manner. However when A and F are non-singular so is AF and we may write $|I + AF| = |A|\,|F|\,|I + (AF)^{-1}|$ and

$$\Delta I(A) = \tfrac{1}{2}\log|A| + \tfrac{1}{2}\log|F| + \tfrac{1}{2}\log|I + (AF)^{-1}|.$$

If the elements of $(AF)^{-1}$ are small we have

$$\Delta I(A) \cong \tfrac{1}{2}\log|A| + \tfrac{1}{2}\log|F| \text{ and } \Delta I_1(A) - \Delta I_2(A) \cong \tfrac{1}{2}\log(|F_1| / |F_2|).$$

So we obtain the criterion $|F|$. The conditions under which it is valid are when, roughly speaking, either

(i) all the diagonal elements of A are large representing large prior uncertainity for all the parameters or

(ii) all the diagonal elements of F are large which is usually so if the n of (1.1) is large.

We now introduce a criterion based on the Neyman-Pearson theory of tests and show that a particular case of it leads to the $|F|$ criterion. Lehmann has given [6] a proof that for the experiment (1.1) the uniformly-most-powerful invariant test of the hypothesis $H_0: \theta = 0$ is provided by the usual $\mathfrak{F}$-test based on

$$\mathfrak{F} = \frac{\hat{\theta}' F \hat{\theta}/k}{(y - X\hat{\theta})C^{-1}(y - X\hat{\theta})/(n - k)}$$

where $\hat{\theta}$ are the maximum-likelihood estimates of θ. Taking as critical region $\mathfrak{F} > \mathfrak{F}_0$, denote by $P_{II}(\theta)$ the probability of error of the second kind under the alternative hypothesis $H: \theta$. A criterion for design can be stated as follows: "Take a probability density function for θ, $p(\theta)$, and choose the experiment to minimise $\int p(\theta)P_{II}(\theta)\,d\theta$." The choice of $p(\theta)$ is arbitrary but in a situation where we are initially very uncertain about θ it would be sensible to take $p(\theta)$ to be normal with mean 0 and diagonal dispersion matrix with variances all equal to E and consider what happens as $E \to \infty$. This we now do and state the theorem:

THEOREM 5.1. *If $p(\theta/E)$ is the probability density function just described then choosing the experiment to minimise $\int p(\theta/E)P_{II}(\theta)\,d\theta$ is equivalent to choosing it to maximise $|F|$.*

PROOF. Tang has shown [8] that $P_{II}(\theta)$ depends solely on the function $\lambda = \frac{1}{2}\theta' F \theta$. In fact

$$P_{II}(\theta) = \sum_{i=0}^{\infty} c_i \lambda^i e^{-\lambda}/i!$$

where the c_i are functions of i, $\mathfrak{F}_0$, k, n but are independent of X and C and also $c_i \to 0$ as $i \to \infty$ thus making the series uniformly convergent in $0 \leqq \lambda < \infty$.

$$\int p(\theta/E)P_{II}(\theta)\,d\theta = (2\pi E)^{-\frac{1}{2}} \int \exp(-\tfrac{1}{2}\theta' E^{-1}\theta)P_{II}(\theta)\,d\theta$$

and

$$\lim_{E\to\infty} \int \exp\left(-\tfrac{1}{2}\theta' E^{-1}\theta\right) P_{II}(\theta)\, d\theta = \int P_{II}(\theta)\, d\theta.$$

Therefore

$$\int p(\theta/E) P_{II}(\theta)\, d\theta \sim (2\pi E)^{-\frac{1}{2}k} \int P_{II}(\theta)\, d\theta = (2\pi E)^{-\frac{1}{2}k} \int \left(\sum_{i=0}^{\infty} c_i \lambda^i e^{-\lambda}/i! \right) d\theta$$

$$= (2\pi E)^{-\frac{1}{2}k} \sum_{i=0}^{\infty} c_i \int \lambda^i e^{-\lambda}\, d\theta/i!$$

since the series is uniformly convergent in $0 \leqq \lambda < \infty$. Now

$$\int \lambda^i e^{-\lambda}\, d\theta = 2^{-i} \int (\theta' F \theta)^i \exp\left(-\tfrac{1}{2}\theta' F \theta\right) d\theta$$

$$= \frac{(2\pi)^{\frac{1}{2}k}}{2^i |F|^{\frac{1}{2}}} \int \frac{|F|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}k}} \exp\left(-\tfrac{1}{2}\theta' F \theta\right)(\theta' F \theta)^i\, d\theta.$$

Now under the probability density function $(2\pi)^{-\frac{1}{2}k}|F|^{\frac{1}{2}} \exp\left(-\tfrac{1}{2}\theta' F \theta\right)$, $\theta' F \theta$ is distributed as chi-square with k degrees of freedom. Hence

$$\int (2\pi)^{-\frac{1}{2}k} |F|^{\frac{1}{2}} \exp\left(-\tfrac{1}{2}\theta' F \theta\right)(\theta' F \theta)^i\, d\theta = g(i, k)$$

say where $g(i, k)$ is a function of i and k only. Therefore

$$\int p(\theta/E) P_{II}(\theta)\, d\theta \sim E^{-\frac{1}{2}k}\left(\sum c_i g(i, k)/i!\, 2^i\right) |F|^{-\frac{1}{2}} \propto |F|^{-\frac{1}{2}}.$$

Hence minimising $\int p(\theta/E) P_{II}(\theta)\, d\theta$ is equivalent to maximising $|F|$.

We give now an example of the use of the $|F|$ criterion, which has been treated by Tocher [9] from another viewpoint. If $C = I$ then $F = X'X$. Suppose the allocation vectors $x_i = (X_{i1}, \cdots, X_{ik})$ $(i = 1, \cdots, n)$ can be varied subject to the restriction $\sum_{i=1}^{n} X_{ij}^2 = a_j$. Then:

THEOREM 5.2. *If $\sum_{i=1}^{n} X_{ij}^2 = a_j$ $(j = 1, \cdots, k)$ where $a_1, \cdots, a_k$ are positive constants then $|F|$ is maximum when $x_1, \cdots, x_n$ are chosen so that $F_{rs} = 0$, $r \neq s$, i.e. when the design is orthogonal.*

PROOF.

$$F_{rs} = \sum_{i=1}^{n} X_{ir} X_{is}$$

Therefore, by a well-known property of positive-definite matrices

$$|F| \leqq \left(\sum X_{i1}^2\right)\left(\sum X_{i2}^2\right) \cdots \left(\sum X_{ik}^2\right) = a_1 a_2 \cdots a_k.$$

But when $F_{rs} = 0$, $r \neq s$, $|F| = a_1 a_2 \cdots a_k$. Therefore $|F|$ is maximum when the design is orthogonal.

6. We now consider the modifications in the $|F|$ criterion imposed by the presence of nuisance-parameters, ϕ, which enter linearly into the expressions for the expectations of our random variables, Y, just as the parameters of interest, θ, do.

Let there be q nuisance parameters and suppose that Y is now normal with mean $X\theta + Z\phi$ and dispersion matrix I. For simplicity take the case where

$$F_1 = \begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z \end{pmatrix}$$

is non-singular. Then θ and ϕ are estimable by the maximum-likelihood estimates $\hat{\theta}$ and $\hat{\phi}$. Write

$$\omega = \begin{pmatrix} \theta \\ \phi \end{pmatrix} \quad \text{and} \quad \hat{\omega} = \begin{pmatrix} \hat{\theta} \\ \hat{\phi} \end{pmatrix}.$$

Then

$$p(\hat{\omega}/\omega) = (2\pi)^{-\frac{1}{2}(k+q)} |F_1|^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\hat{\omega} - \omega)'F_1(\hat{\omega} - \omega)\right].$$

Suppose that the prior distribution for ω is

$$p(\omega) = (2\pi)^{-\frac{1}{2}(k+q)} |D|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\omega - \omega_0)D^{-1}(\omega - \omega_0)\right]$$

where

$$D = \begin{pmatrix} A & E \\ E' & B \end{pmatrix}$$

(A and B are the prior dispersion matrices for θ and ϕ respectively.) By Bayes' Theorem we find that $p(\omega/\hat{\omega})$ is normal with information matrix $(F_1 + D^{-1})$. To find the information about θ we must integrate out ϕ in (a) $p(\omega)$ and (b) $p(\omega/\hat{\omega})$ to obtain the marginal distributions of θ. We find:

(a) $p(\theta)$ is normal with dispersion matrix A

(b) $p(\theta/\hat{\omega})$ is normal with dispersion matrix L where L is the leading $k \times k$ diagonal sub-matrix of $(F_1 + D^{-1})^{-1}$. Then

$$Ip(\theta) = -\tfrac{1}{2} \log\left[(2\pi)^{\frac{1}{2}k} |A|\right] - \tfrac{1}{2}k$$

and

$$Ip(\theta/\hat{\omega}) = -\tfrac{1}{2} \log\left[(2\pi)^{\frac{1}{2}k} |L|\right] - \tfrac{1}{2}k$$

and

$$\Delta I = Ip(\theta/\hat{\omega}) - Ip(\theta) = -\tfrac{1}{2} \log |L| + \tfrac{1}{2} \log |A|. \tag{6.1}$$

If the elements of $(DF_1)^{-1}$ are small

$$(F_1 + D^{-1})^{-1} = F_1^{-1}(I + (DF_1)^{-1})^{-1} \cong F_1^{-1}.$$

Write

$$F_1^{-1} = \begin{pmatrix} \mathfrak{A} & \gamma \\ \gamma' & \beta \end{pmatrix} \tag{6.2}$$

where $\mathfrak{A}$ is the $k \times k$ dispersion matrix of $\hat{\theta}$ in $p(\hat{\omega}/\omega)$. Then $L \cong \mathfrak{A}$ and

$$\Delta I = -\tfrac{1}{2} \log |\mathfrak{A}| + \tfrac{1}{2} \log |A|. \tag{6.3}$$

The conditions under which the elements of $(DF_1)^{-1}$ are small are when, roughly speaking, either

(i) all the diagonal elements of D are large corresponding to large prior ignorance of all the parameters or

(ii) all the diagonal elements of F_1 are large corresponding to a "strong" experiment.

So, by (6.3), we see that under the conditions stated maximising ΔI is equivalent to minimising $|\mathfrak{A}|$.

A. Wald [10] developed the use of $|\mathfrak{A}|$ which he called the "generalised variance" but his justification of it was pragmatical rather than logical.

In most problems it is usually a simple matter to calculate F_1 from the allocation vectors. Then by Jacobi's Theorem we obtain $|\mathfrak{A}| = |Z'Z| \,/\, |F_1|$.

EXAMPLE 1. A simple 2^2 factorial experiment without interaction with the base level (both factors absent) as the nuisance parameter. $k = 2$; $q = 1$.

$$\begin{matrix} & n_0 & n_1 & n_2 & n_3 \\ X' = & \begin{pmatrix} 0 \cdots 0 & 1 \cdots 1 & 0 \cdots 0 & 1 \cdots 1 \\ 0 \cdots 0 & 0 \cdots 0 & 1 \cdots 1 & 1 \cdots 1 \end{pmatrix} \\ Z' = & (1 \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots 1) \end{matrix}$$

Suppose $n = n_0 + n_1 + n_2 + n_3 = 4m$. Then

$$F_1 = \begin{pmatrix} n_1 + n_3 & n_3 & n_1 + n_3 \\ n_3 & n_2 + n_3 & n_2 + n_3 \\ n_1 + n_3 & n_2 + n_3 & n \end{pmatrix}$$

$$|\mathfrak{A}| = n/|F_1| = n/(n_1n_2n_3 + n_0n_2n_3 + n_0n_1n_3 + n_0n_1n_2).$$

For minimum $|\mathfrak{A}|$, $n_i = m$ $(i = 0, 1, 2, 3)$.

EXAMPLE 2. The addition of an interaction term θ_3 to Example 1.

$$\begin{matrix} & n_0 & n_1 & n_2 & n_3 \\ X' = & \begin{pmatrix} 0 \cdots 0 & 1 \cdots 1 & 0 \cdots 0 & 1 \cdots 1 \\ 0 \cdots 0 & 0 \cdots 0 & 1 \cdots 1 & 1 \cdots 1 \\ 0 \cdots 0 & 0 \cdots 0 & 0 \cdots 0 & 1 \cdots 1 \end{pmatrix} \\ Z' = & (1 \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots 1) \end{matrix}$$

Suppose $n = 4m$. Then

$$F_1 = \begin{pmatrix} n_1 + n_3 & n_3 & n_3 & n_1 + n_3 \\ n_3 & n_2 + n_3 & n_3 & n_2 + n_3 \\ n_3 & n_3 & n_3 & n_3 \\ n_1 + n_3 & n_2 + n_3 & n_3 & n \end{pmatrix}$$

$$|\mathfrak{A}| = n/n_0 n_1 n_2 n_3 .$$

For minimum $|\mathfrak{A}|$, $n_i = m$ $(i = 0, 1, 2, 3)$.

EXAMPLE 3. k treatments and a control [3]; $q = 1$.

$$X' = \begin{array}{c} \begin{matrix} n_0 & & n_1 & & n_2 & & & n_k \end{matrix} \\ \begin{pmatrix} 0 \cdots 0 & 1 \cdots 1 & 0 \cdots 0 & \cdots & 0 \cdots 0 \\ 0 \cdots 0 & 0 \cdots 0 & 1 \cdots 1 & \cdots & 0 \cdots 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 \cdots 0 & 0 \cdots 0 & 0 \cdots 0 & & 1 \cdots 1 \end{pmatrix} \end{array}$$

$$Z' = (1 \cdots 1)$$

Suppose n is divisible by $k + 1$. Then

$$F_1 = \begin{pmatrix} n_1 & & & n_1 \\ & \ddots & 0 & \vdots \\ 0 & & n_k & n_k \\ n_1 & \cdots & n_k & n \end{pmatrix}$$

$$|\mathfrak{A}| = n/|F_1| = n/n_0 n_1 \cdots n_k$$

which is minimised when $n_i = n/(k + 1)$ $(i = 0, 1, \cdots, k)$.

The calculation of $|\mathfrak{A}| = |Z'Z| \,/\, |F_1|$, though simple in the examples given, may be complicated. Then it may be possible to use a method we now elaborate.

DEFINITION. An experiment involving nuisance-parameters is "part-orthogonal" if $\gamma = 0$. [See (6.2).]

The customary definition of "orthogonal" requires that all non-diagonal elements of F_1^{-1} should be zero. However, if $\gamma = 0$, we could achieve this condition by two separate orthogonal transformations of θ and ϕ. From an informational point of view these are irrelevant.

By Wegner's Theorem $|F_1^{-1}| \leqq |\alpha| \, |\beta|$ with equality when $\gamma = 0$. But by Jacobi's Theorem $|X'X| = |\beta| \,/\, |F_1^{-1}|$. Hence $|\alpha| \geqq 1/|X'X|$ with equality when $\gamma = 0$. From this we derive the working principle: "Find the design which maximises $|X'X|$; if this is a part-orthogonal design (experiment) then it is the design which minimises $|\mathfrak{A}|$."

The allocation problem of Section 3 remains important in the presence of nuisance parameters. We prove a theorem which generalises Thorem 3.1. It is analogous to one due to Chernoff [2]. Now, the allocation vectors are the rows of the $n \times (k + q)$ matrix (XZ); denote them by u. Given g possible allocation

vectors $u(i) = \begin{pmatrix} x(i) \\ z(i) \end{pmatrix}$ $(i = 1, \cdots, g)$ denote by $\Delta I(n_1, \cdots n_g)$ the information about θ in the experiment consisting of n_i observations at $u(i)$ $(i = 1, \cdots, g)$ with the errors (η) uncorrelated with unit variance.

THEOREM 6.1.[1] *When n is large, $\Delta I(n_1, \cdots, n_g)$ is in general maximised when no more than $\frac{1}{2}k(k + 1 + 2q)$ of the allocation vectors are used.* (It is not necessary that F_1 be non-singular.)

PROOF. By (6.1), $\Delta I(n_1, \cdots, n_g) = -\frac{1}{2} \log |L(n_1, \cdots, n_g)| + \frac{1}{2} \log |A|$ where $L(n_1, \cdots, n_g)$ is the leading $k \times k$ diagonal sub-matrix of $(F_1 + D^{-1})^{-1}$ with $F_1 = F_1(n_1, \cdots, n_g)$ and $[F_1]_{rs} = \sum_{i=1}^{g} n_i u_r(i) u_s(i)$. Suppose

$$D^{-1} = \begin{pmatrix} P & S \\ S' & R \end{pmatrix}$$

where P is $k \times k$. Then by Jacobi's Theorem

$$|L(n_1, \cdots, n_g)| = |Z'Z + R| \,/\, |F_1 + D^{-1}|.$$

Replace n_i by np_i, $p_i \geqq 0$, $\sum p_i = 1$:

$$L(n_1, \cdots, n_g) \to L[np]$$

$$F_1 \to F_1[np]$$

where

$$F_1[np] = \begin{pmatrix} F[np] & G[np] \\ G'[np] & H[np] \end{pmatrix}$$

say

$$Z'Z \to H[np]$$

$$|L[np]| = |H[np] + R| \,/\, |F_1[np] + D^{-1}|$$

We show that $|L[np]|$ is minimised when no more than $\frac{1}{2}k(k + 1 + 2q)$ of the p_i's are non-zero:

(a) If $|L[np]|$ is minimum at $p = p(m)$ where $p_i(m) = 1$ then, since $1 \leqq \frac{1}{2}k(k + 1 + 2q)$, the statement holds.

(b) If $|L[np]|$ is minimum at $p = p(m)$ when more than one $p_i(m)$ 0, $\neq$ we proceed as follows.

$$\frac{\partial}{\partial p_i} \log |L[np]| = \frac{1}{|H[np] + R|} \frac{\partial}{\partial p_i} |H[np] + R|$$

$$- \frac{1}{|F_1[np] + D^{-1}|} \frac{\partial}{\partial p_i} |F_1[np] + D^{-1}|.$$

But

$$H[np]_{rs} = \sum_{i=1}^{g} np_i z_r(i) z_s(i)$$

$$F_1[np]_{rs} = \sum_{i=1}^{g} np_i u_r(i) u_s(i)$$

[1] The author is indebted to a referee for suggesting this theorem.

and by row-by-row differentiation of the determinants we find

$$\frac{\partial}{\partial p_i} | H[np] + R | = nz'(i) \text{ adj } (H[np] + R)z(i)$$

$$\frac{\partial}{\partial p_i} | F_1[np] + D^{-1} | = nu'(i) \text{ adj } (F_1[np] + D^{-1})u(i).$$

Therefore

$$\frac{\partial}{\partial p_i} \log | L[np] | = n [z'(i)(H[np] + R)^{-1}z(i) - u'(i)(F_1[np] + D^{-1})^{-1}u(i)] \tag{6.4}$$

$$= -nu'(i)Q[np]u(i)$$

where

$$Q[np] = (F_1[np] + D^{-1})^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & (H[np] + R)^{-1} \end{pmatrix}$$

$$= (F_1[np] + D^{-1})^{-1} \begin{pmatrix} I & -(G[np] + S)(H[np] + R)^{-1} \\ 0 & 0 \end{pmatrix}$$

Therefore the rank of $Q[np]$ is k. But $\sum_1^q p_i = 1$, therefore by the method of undetermined multipliers applied to variations of the p_i's for i's for which $p_i(m) \neq 0$, we see that for such i there exists a λ such that

$$\left(\frac{\partial}{\partial p_i} \log | L[np] |\right)_{p(m)} = \lambda$$

or, by (6.4), $u'(i)Q[np(m)]\, u(i) = -\lambda/n$. Hence those allocation vectors which are used to minimise $|L[np]|$ must lie on a central quadric of rank k. Such a quadric is determined by $(k + q) + (k + q - 1) + \cdots + (q + 1) = \frac{1}{2}k(k + 1 + 2q)$ constants implying that in general no more than $\frac{1}{2}k(k + 1 + 2q)$ of the vectors can have their associated $p_i(m)$'s non-zero. Now

$$\Delta I(n_1, \cdots, n_q) = -\tfrac{1}{2} (\log |L[np]|)_{p_i = n_i/n} + \tfrac{1}{2} \log |A|.$$

Therefore, when n is large, we can say that in general $\Delta I(n_1, \cdots, n_q)$ may be maximised when no more than $\frac{1}{2}k(k + 1 + 2q)$ of the n_i are non-zero and the theorem is proved.

REFERENCES

[1] G. E. P. Box and J. S. Hunter, "Multifactor experimental designs for exploring response surfaces," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 195–241.

[2] H. Chernoff, "Locally optimum designs for estimating parameters," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 586–602.

[3] C. W. Dunnett, "Multiple comparison procedures for comparing several treatments with control," *J. Amer. Stat. Assn.*, Vol. 50 (1955), pp. 1096–1121.

[4] S. Ehrenfeld, "Complete class theorems in experimental designs," *Third Berkeley Symposium*, Vol. 1, pp. 57–67.

[5] G. Elfving, "Optimum allocation in regression theory," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 255–262.

[6] E. L. LEHMANN, "Lecture notes," University of California.
[7] D. V. LINDLEY, "On a measure of the information provided by an experiment," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 986–1005.
[8] P. C. TANG, "The power function of the analysis of variance," *Stat. Res. Memoirs*, Vol. 2, pp. 126–149.
[9] K. D. TOCHER, "A note on the design problem," *Biometrika*, Vol. 39 (1952), p. 189.
[10] A. WALD, "On the efficient design of statistical investigations," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 134–140.

NOTE ON ESTIMATING INFORMATION[1]

By Colin R. Blyth

University of Illinois and *Stanford University*

1. Summary. This note is concerned with estimation of the Shannon-Wiener measure of information. Low bias estimates are obtained, and their bias and variance. These estimates are extended to the case where the number of possible values of the random variable is not known. The estimates are compared asymptotically with the maximum likelihood estimates. They are also compared with the minimax estimates (for squared error loss function) for a few special cases where these are easily found.

2. Introduction. Consider a random variable Y with finitely many distinct possible values:

$$P(Y = a_i) = p_i, \qquad i = 1, \cdots, k.$$

A *metric* measure of dispersion of Y measures how spread out the distribution of Y is, in terms of distance in the space of Y. If this space has no relevant distance function (e.g., $k = 3$, a_1 = green, a_2 = red, a_3 = white), there is no relevant metric measure of dispersion. An *absolute* measure of dispersion of Y measures the degree to which the total probability of 1 is broken up into pieces in the distribution of Y. Such a measure is a function of $p_1, \cdots, p_k$ only; is free from dependence on the a_i's; is large when the probability is much broken up (e.g., $p_1, \cdots, p_k = 1/k, \cdots, 1/k$), small when it is not much broken up (e.g., $p_1, \cdots, p_k = .99, .01, 0, \cdots, 0$).

In handling both kinds of dispersion measures the following addition property plays the same important role: {Divide the values of Y into groups. Dispersion of Y = between group dispersion + expected within group dispersion.}. Knowing the distribution of Y gives information useful in predicting Y. Actually observing Y gives additional information—enough for perfect prediction. This additional information can be called information in Y or unpredictability of Y and can be measured by a measure of dispersion. In this language the addition property says that the information in observing Y equals the information in observing which group Y falls in plus the expected information in observing which member of that group.

For real valued Y the addition property identifies variance (except for a constant multiplier) among all metric dispersion measures of the form $Ef(|Y - EY|)$ with f continuous. This easily extends to weighted averages of the partial variances when Y has values in a Euclidean n-space. Similarly the addition property identifies information or entropy $H = -\sum p_i \log_2 p_i$ (except for a constant multiplier) among all absolute dispersion measures $f_k(p_1, \cdots, p_k)$ with

Received January 23, 1958; revised April 18, 1958.

[1] Work supported by the Office of Naval Research.

f_k continuous and $f_k(1/k, \cdots, 1/k)$ an increasing function of k, as is proved in [1].

The addition property leads to very convenient mathematical simplifications. For this reason variance and information H are very widely used dispersion measures. But there seems to be little intuitive necessity for the addition property. Thus in the metric case it seems quite reasonable to use measures like $E|Y - EY|$ which lack the property- and loss functions other than squared error. Similarly in the absolute case it would seem quite reasonable to use measures like the natural chi-square measure $(k-1) - k\sum(p_i - 1/k)^2$ which lack the property. Essentially equivalent to this chi-square measure is the following linear function of it, which is the terms of order up to 2 in a Taylor series for H:

$$H_2 = 1 - \frac{C}{2}\left\{k - 2 + \sum_{i=1}^{k}(2p_i - 1)^2/2\right\}.$$

If $f(1, 0, \cdots, 0) = 0$ is desired we could make the necessary subtraction from the measure H_2.

This note is concerned with estimation of information or entropy of Y:

$$H = H(p_1, \cdots, p_k) = -C\sum_{i=1}^{k} p_i \log p_i,$$

where $C = \log_2 e = 1.442695$ and $p_i \log p_i$ is taken to be 0 whenever $p_i = 0$. Our estimate is to be based on independent repetitions $Y_1, \cdots, Y_n$ of the experiment Y. Then $X_1, \cdots, X_k$, where X_i is the number of Y's with the value a_i, is a sufficient statistic for $p_1, \cdots, p_k$ and has the following multinomial family of possible distributions:

$$P(X_1, \cdots, X_k = x_1, \cdots, x_k) = n!\prod_{i=1}^{k} p_i^{x_i}/x_i!,$$

(1)

$$x_i = 0, 1, \cdots, n, \quad \sum_{i=1}^{k} x_i = n, \quad 0 \leqq p_i \leqq 1, \quad \sum_{i=1}^{k} p_i = 1.$$

We are now concerned, then, with the problem of what function $f_k(X_1, \cdots, X_k)$ to use as an estimate for H. The maximum likelihood estimate is considered by Miller and Madow [2]; it is good when n is large, but is likely to be poor for n small. One reasonable estimate would be the best unbiased one. Upon noticing that there is no unbiased estimate we will consider instead best estimates with low bias.

3. Low bias estimation, k known. Since (1) is a complete family of distributions, the problem of unbiased estimation of any function $g(p_1, \cdots, p_k)$ is solved by Lehmann and Scheffe [3]. In fact, since $Ef(X_1, \cdots, X_k)$ is for every every function f a polynomial in $p_1, \cdots, p_k$ of degree at most n, no functions of $p_1, \cdots, p_k$ other than such polynomials possess unbiased estimates. And using the usual factorial notation $x^{(\nu)} = x(x-1)\cdots(x-\nu+1)$ we have

$$E\{X_1^{(\nu_1)} \cdots X_k^{(\nu_k)}\} = n^{(\nu_1+\cdots+\nu_k)}p_1^{\nu_1} \cdots p_k^{\nu_k},$$

which reduces to $0 \equiv 0$ whenever $\sum_{i=1}^{k} \nu_i > n$ but not otherwise. We therefore have

$$E \sum c(\nu_1, \cdots, \nu_k) X_1^{(\nu_1)} \cdots X_k^{(\nu_k)} / n^{(\nu_1+\cdots+\nu_k)} = \sum c(\nu_1, \cdots, \nu_k) p_1^{\nu_1} \cdots p_k^{\nu_k},$$

where summation is over any set of $(\nu_1, \cdots, \nu_k)$'s with $\sum_{i=1}^{k} \nu_i \leqq n$ for every member. It follows from the completeness that $\sum c(\nu_1, \cdots, \nu_k) X_1^{(\nu_1)} \cdots X_k^{(\nu_k)} / n^{(\nu_1+\cdots+\nu_k)}$ is the unique uniformly minimum variance (U.M.V.) unbiased estimate of $\sum c(\nu_1, \cdots, \nu_k) p_1^{\nu_1} \cdots p_k^{\nu_k}$ whenever $\sum_{i=1}^{k} \nu_i \leqq n$ for every term of the sum. This solves the problem of unbiased estimation of all functions $g(p_1, \cdots, p_k)$ because the U.M.V. unbiased estimate of every degree $\leqq n$ polynomial in $p_1, \cdots, p_k$ has been written down and no other functions of $p_1, \cdots, p_k$ possess unbiased estimates.

It is now clear that there is no unbiased estimate of H. If low bias is what we want the next best thing would be to use the U.M.V. unbiased estimate of the degree n polynomial which is in some sense (smallest maximum distance apart, for example) closest to H. A much more easily obtained polynomial which agrees quite closely with H is the terms of degree $\leqq n$ in the Taylor series expansion of H about the point $(\frac{1}{2}, \cdots, \frac{1}{2})$. We will consider use of the U.M.V. unbiased estimate of this polynomial as an estimate for H.

Writing $\gamma_i = p_i - \frac{1}{2}$ we have

$$\begin{aligned} p_i \log p_i &= (\tfrac{1}{2} + \gamma_i) \log (\tfrac{1}{2} + \gamma_i) \\ &= -\frac{1}{2C} + \left(1 - \frac{1}{C}\right)\gamma_i + \frac{1}{2}\left\{\frac{(2\gamma_i)^2}{1 \cdot 2} - \frac{(2\gamma_i)^3}{2 \cdot 3} + \frac{(2\gamma_i)^4}{3 \cdot 4} - \cdots\right\}. \end{aligned}$$

Hence

$$\begin{aligned} H &= -C \sum_{i=1}^{k} p_i \log p_i \\ &= 1 - \frac{C}{2}\left\{\sum_{i=1}^{k} (2\gamma_i) + \sum_{i=1}^{k} \frac{(2\gamma_i)^2}{1 \cdot 2} - \sum_{i=1}^{k} \frac{(2\gamma_i)^3}{2 \cdot 3} + \cdots\right\} \\ &= 1 - \frac{C}{2}\left\{(2 - k) + \sum_{\alpha=2}^{\infty} \sum_{i=1}^{k} (-2\gamma_i)^\alpha / \alpha^{(2)}\right\}. \end{aligned}$$

Here $|\, 2\gamma_i \,| \leqq 1$ so all series converge absolutely and can be rearranged. Also, $\sum_{i=1}^{k} (2\gamma_i) = 2 - k$. For any integer $r \leqq n$ we now write

$$\begin{aligned} H_r &= 1 - \frac{C}{2}\left\{(2 - k) + \sum_{\alpha=2}^{r} \sum_{i=1}^{k} (-2\gamma_i)^\alpha / \alpha^{(2)}\right\} \\ &= 1 - \frac{C}{2}\left\{(2 - k) + \sum_{\alpha=2}^{r} \sum_{i=1}^{k} \frac{1}{\alpha^{(2)}} \sum_{\nu=0}^{\alpha} \binom{\alpha}{\nu} (-2p_i)^\nu\right\}. \end{aligned}$$

The U.M.V. unbiased estimate of H_r is

$$Z_r = 1 - \frac{C}{2}\left\{(2 - k) + \sum_{\alpha=2}^{r} \sum_{i=1}^{k} \sum_{\nu=0}^{\alpha} \frac{1}{\alpha^{(2)}} \binom{\alpha}{\nu} (-2)^\nu \frac{X_i^{(\nu)}}{n^{(\nu)}}\right\}.$$

The bias of Z_r as an estimate for H is

$$B_r = EZ_r - H = H_r - H$$

$$= \frac{C}{2} \sum_{\alpha=r+1}^{\infty} \sum_{i=1}^{k} (-2\gamma_i)^\alpha / \alpha^{(2)}.$$

Now u^s is a convex function of u for s an even integer, and $-u^s$ is convex on $u \leqq 0$ for s an odd integer. From this it is easily shown that if $\sum_{i=1}^{k} u_i = 2 - k$ and $|u_i| \leqq 1$, then

$$(k - 1 + (-1)^s)\left(1 - \frac{2}{k}\right)^s \leqq (-1)^s \sum_{i=1}^{k} u_i^s \leqq k - 1 + (-1)^s.$$

These lower, upper bounds are achieved by the choices $(u_1, \cdots, u_k) = (2/k - 1, \cdots, 2/k - 1)$ and $(1, -1, \cdots, -1)$ respectively. Applying this to the series for B_r gives

$$\frac{C}{2} \sum_{\alpha=r+1}^{\infty} \frac{(k - 1 + (-1)^\alpha)(1 - 2/k)^\alpha}{\alpha^{(2)}} \leqq B_r \leqq \frac{C}{2}\left\{\frac{k - 1 - (-1)^r}{r} - 2 \sum_{\alpha=r+1}^{\infty} \frac{(-1)^\alpha}{\alpha}\right\}.$$

This lower bound is achieved when the p_i's are all $1/k$, and the upper bound is achieved when some $p_i = 1$. If $k > 2$ this lower bound is positive and we will use the estimate

$$Z_r' = Z_r - \frac{C}{2} \sum_{\alpha=r+1}^{\infty} \frac{(k - 1 + (-1)^\alpha)(1 - 2/k)^\alpha}{\alpha^{(2)}}$$

instead of Z_r for H because Z_r' has the same variance as Z_r and uniformly smaller bias. For a fixed set $\gamma_1, \cdots, \gamma_k$ we have

$$\bar{B}_r \leqq \frac{Ck}{2r} (\max_i |2\gamma_i|)^{r+1}$$

and the corresponding result for the bias of the improved estimate Z_r'. To compute the variance of Z_r we now use the fact

$$X_i^{(\nu_1)} = \sum_{t=0}^{\nu_1} \frac{\nu_1^{(t)} \nu_2^{(t)}}{t!} (X_i - \nu_2)^{(\nu_1 - t)},$$

which gives

$$X_i^{(\nu_1)} X_i^{(\nu_2)} = \sum_{t=0}^{\min(\nu_1, \nu_2)} \frac{\nu_1^{(t)} \nu_2^{(t)}}{t!} X_i^{(\nu_1 + \nu_2 - t)}.$$

Further routine calculations now give

$$E\left(\frac{X_i^{(\nu_1)}}{n^{(\nu_1)}} - p_i^{\nu_1}\right)\left(\frac{X_i^{(\nu_2)}}{n^{(\nu_2)}} - p_i^{\nu_2}\right) = \sum_{t=1}^{\min(\nu_1, \nu_2)} \frac{\nu_1^{(t)} \nu_2^{(t)}}{t!} \frac{p_i^{\nu_1 + \nu_2 - t}(1 - p_i)^t}{n^{(t)}},$$

$$E\left(\frac{X_i^{(\nu_1)}}{n^{(\nu_1)}} - p_i^{\nu_1}\right)\left(\frac{X_j^{(\nu_2)}}{n^{(\nu_2)}} - p_j^{\nu_2}\right) = \left(\frac{n^{(\nu_1 + \nu_2)}}{n^{(\nu_1)} n^{(\nu_2)}} - 1\right) p_i^{(\nu_1)} p_j^{(\nu_2)}, \qquad i \neq j.$$

From these, the variance of Z_r is seen to be

$$\begin{aligned}\operatorname{var} Z_r &= E(Z_r - H_r)^2 \\ &= \frac{C^2}{4} E\left\{\sum_{\alpha=2}^{r}\sum_{\nu=0}^{\alpha}\binom{\alpha}{\nu}\frac{(-2)^\nu}{\alpha^{(2)}}\sum_{i=1}^{k}\left(\frac{X_i^{(\nu)}}{n^{(\nu)}} - p_i^\nu\right)\right\}^2 \\ &= \frac{C^2}{4}\sum_{\alpha_1=2}^{r}\sum_{\alpha_2=2}^{r}\sum_{\nu_1=0}^{\alpha_1}\sum_{\nu_2=0}^{\alpha_2}\binom{\alpha_1}{\nu_1}\binom{\alpha_2}{\nu_2}\frac{(-2)^{\nu_1+\nu_2}}{\alpha_1^{(2)}\alpha_2^{(2)}}\left\{\sum_{i_1\neq i_2=1}^{k} p_{i_1}^{\nu_1} p_{i_2}^{\nu_2}\left[\frac{n^{(\nu_1+\nu_2)}}{n^{(\nu_1)}n^{(\nu_2)}} - 1\right]\right. \\ &\qquad \left. + \sum_{i=1}^{k}\sum_{t=1}^{\min(\nu_1,\nu_2)}\frac{\nu_1^{(t)}\nu_2^{(t)}}{t!}\frac{p_i^{\nu_1+\nu_2-t}(1-p_i)^t}{n^{(t)}}\right\}.\end{aligned}$$

Grouping together terms of like order in n, the asymptotic variance as $n \to \infty$ of the sequence $\{Z_n\}$ of estimates is seen to be

$$\operatorname{var}(Z_n) \sim \frac{C^2}{n}\sum_{i=1}^{k} p_i(\log p_i - \sum_{j=1}^{k} p_j \log p_j)^2$$

provided the non-zero p_i's are not all equal; and

$$\operatorname{var}(Z_n) \sim \frac{C^2}{n}\cdot\frac{k^* - 1}{2(n-1)}$$

if the non-zero p_i's are all equal, where k^* is the number of non-zero p_i's. And for bias of this sequence of estimates we have asymptotically as $n \to \infty$

$$\frac{kC}{2n(n+1)}\left(1 - \frac{2}{k}\right)^{n+2} \leqq B_n \leqq \frac{kC}{2n}\{\max_i |2\gamma_i|\}^{n+1}.$$

Comparison with asymptotic results obtained by Miller and Madow for the maximum likelihood estimates $\{H_n'\}$ and the estimates $\{H_n''\} = \{H_n' + C(k-1)/2n\}$ shows the following: No asymptotic differences in variance. Asymptotic differences in expected square error only in the special cases where this has order $1/n^2$ or smaller. Bias is asymptotically much smaller for $\{Z_n\}$ than for $\{H_n'\}$ or $\{H_n''\}$ except for the special case when some $p_i = 1$.

For small n, numerical checking shows the following: Z_n has a smaller bias than H_n' or H_n'' over most of the range of $(p_1, \cdots, p_k)$. Comparing expected squared error as a whole, H_n' is quite poor and there is little to choose between Z_n and H_n'': sometimes one seems better, sometimes the other. In these comparisons Z_n and H_n'' are modified by substituting $C \log k$ for any value exceeding $C \log k$, since this uniformly reduces expected squared error. For example in Table 1 for $k = 2$, every value exceeding $C \log 2 = 1$ would be replaced by 1. This table gives $Z_n(x_1)$ in the upper part of each column and $H_n''(x_1)$ in the lower part, for $n = 2, 3, \cdots, 7$ and all possible x_1. Values not tabled are obtained from $Z_n(n - x_1) = Z_n(x_1)$ and $H_n''(n - x_1) = H_n''(x_1)$.

When $k = 2$, minimax estimates (squared error loss function can be found for small n by the usual method of guessing a least favorable a priori distribution λ for $p_1 = p$ and finding the corresponding Bayes estimate Z_λ. If the risk function

TABLE 1

Estimates of H for $k=2$ with $Z_n(x_1)$ in upper part of each column, $H''_n(x_1)$ in lower part

n / x_1	2	3	4	5	6	7
0	.27865	.27865	.15843	.15843	.11034	.11034
1	1.72135 1.36067	1.24045	1.12022	.92787	.84772	.74238
2	.36067	1.15875	1.12022 1.18034	1.12022	1.00801	.96222
3		.24045	.99162	1.11524	1.08516 1.12022	1.09961
4			.18034	.86619	1.03852	1.08828
5				.14427	.77024	.96617
6					.12022	.69472
7						.10305

TABLE 2

Comparison of low bias Z'_n and minimax Z^* estimates of H for $k=2$

n	$Z^*(0)$, $Z^*(1)$	$\sup_p R_{Z^*}(p)$	$\sup_p R_{Z'_n}(p)$
1	1/2, 1/2	.2500	1
2	$\sqrt{2}-1$, 1	.1716	.2602
3	.33673, .94400	.1134	.1444

$R_{Z_\lambda}(p)$ assumes its maximum value with λ-probability 1 then this λ is indeed least favorable and Z_λ in minimax. Here for λ we take unspecified probabilities at $n+1$ unspecified points one of which is 0, with the restriction that λ be symmetric about $p=\frac{1}{2}$. The points and probabilities are then determined so that $R_{Z_\lambda}(p)$ will have equal maxima at these points. We will compare the risk functions $R_{Z'_n}(p)$ and $R_{Z^*}(p)$ [$Z'_n = Z_n$ except $Z'_n = 1$ when $Z'_n > 1$; Z^* is minimax]. One point of interest is the degree to which $\sup_p R_{Z'_n}(p)$ exceeds $\sup_p R_{Z^*}(p)$. This comparison, given in Table 2 for $n = 1, 2, 3$, is of particular interest for small values of n where Z'_n would be expected to show up most poorly. Actually Z'_n does quite well even for $n = 2, 3$: Z'_n seems to deviate from unbiasedness in the direction of being like the minimax estimate.

Similarly, for $k = 2$, we can compare the minimax and U.M.V. unbiased estimates of the chi square dispersion measure H_2. Equivalently we can compare the minimax estimate T^* and the U.M.V. unbiased estimate T of the binomial variance pq, of which H_2 is just a linear function. This comparison is given for $n = 1, 2, \cdots, 5$ in Table 3. Note that T is very poor compared to T^* for small n, compares more favorably as n increases. For example the ratio $\sup_p R_T(p)/\sup_p R_{T^*}(p)$ is 5.83 when $n = 2$, decreases to 3.37 by $n = 5$. The comparison indicates that for small n the minimax estimate for binomial variance is decidedly

TABLE 3

Comparison of minimax T^* and U.M.V. unbiased T estimates for binomial variance pq

n	$T^*(0), T^*(1), T^*(2)$	$\sup_p R_{T^*}(p)$	$T(0), T(1), T(2)$	$\sup_p R_T(p)$
1	.125, .125, —	.015625	—	—
2	.10355, .25, .10355	.010724	0, 1/2, 0	.062500
3	.08333, .25, .25	.006944	0, 1/3, 1/3	.027778
4	.07158, .20228, .24584	.005124	0, 1/4, 1/3	.018750
5	.06508, .17797, .22841	.004235	0, 1/5, 3/10	.014286

preferable to the U.M.V. unbiased estimate. The estimate T^* is found in the same way as Z^* except that for $n = 3, 4, 5$ T^* has constant risk and can be more easily found by showing that the only constant risk estimate can be Bayes.

4. Low bias estimation, k unknown. When k is unknown we shall consider the estimates obtained by acting as though k were equal to the observed number of different Y values and using the estimates of the preceding section. Now we have

$$Z_r = 1 - \frac{C}{2}\{W_1 + \cdots + W_k\},$$

where

$$W_i = \left(2\frac{X_i}{n} - 1\right) + \sum_{\alpha=2}^{r}\sum_{\nu=0}^{\alpha}\binom{\alpha}{\nu}\frac{(-2)^\nu}{\alpha^{(\alpha)}}\frac{X_i^{(\nu)}}{n^{(\nu)}}.$$

The modification of Z_r just suggested for use in the case k unknown is

$$Z_r^* = 1 - \frac{C}{2}\{W_1^* + \cdots + W_k^*\},$$

where

$$W_i^* = W_i \qquad \text{if } X_i \neq 0,$$
$$= 0 \qquad \text{if } X_i = 0.$$

Since $W_i = -1/r$ when $X_i = 0$ we have

$$P(W_i^* = W_i) = 1 - (1 - p_i)^n,$$
$$P(W_i^* = W_i + 1/r) = (1 - p_i)^n.$$

Hence

$$EW_i^* = EW_i + \frac{1}{r}(1 - p_i)^n,$$

$$EZ_r^* = 1 - \frac{C}{2}\left\{EW_1 + \cdots + EW_k + \frac{1}{r}\sum_{i=1}^{k}(1 - p_i)^n\right\}$$
$$= EZ_r - \frac{C}{2r}\sum_{i=1}^{k}(1 - p_i)^n = H_r - \frac{C}{2r}\sum_{i=1}^{k}(1 - p_i)^n.$$

The bias of Z_r^* as an estimate for H is

$$B_r^* = EZ_r - H = H_r - H - \frac{C}{2r}\sum_{i=1}^{k}(1-p_i)^n$$

$$= B_r - \frac{C}{2r}\sum_{i=1}^{k}(1-p_i)^n.$$

The variance of Z_r^* is found as follows:

$$\text{var } Z_r^* = \frac{C^2}{4}\left\{\sum_{i=1}^{k}\text{var } W_i^* + \sum_{i\neq j-1}^{k}\text{cov } W_i^* W_j^*\right\},$$

$$\text{var } W_i^* = E(W_i^* - EW_i^*)^2$$

$$= [1-(1-p_i)^n]E\left\{(W_i - EW_i) - \frac{1}{r}(1-p_i)^n\right\}^2$$

$$+ (1-p_i)^n E\left\{(W_i - EW_i) + \frac{1}{r} - \frac{1}{r}(1-p_i)^n\right\}^2$$

$$= E(W_i - EW_i)^2 + \frac{1}{r^2}\{(1-p_i)^n - (1-p_i)^{2n}\}$$

$$= \text{var } W_i + \frac{1}{r^2}\{(1-p_i)^n - (1-p_i)^{2n}\}.$$

Furthermore, we have for $i \neq j$

$$P(W_i^*, W_j^* = W_i, W_j) = 1 - (1-p_i)^n - (1-p_j)^n + (1-p_i-p_j)^n,$$

$$P(W_i^*, W_j^* = W_i + 1/r, W_j) = (1-p_i)^n - (1-p_i-p_j)^n,$$

$$P(W_i^*, W_j^* = W_i, W_j + 1/r) = (1-p_j)^n - (1-p_i-p_j)^n,$$

$$P(W_i^*, W_j^* = W_i + 1/r, W_j + 1/r) = (1-p_i-p_j)^n.$$

So the covariance of W_i^*, W_j^*, $i \neq j$, is

$$\text{cov } W_i^*, W_j^* = E(W_i^* - EW_i^*)(W_j^* - EW_j^*)$$

$$= [1-(1-p_i)^n - (1-p_j)^n - (1-p_i-p_j)^n]$$

$$\cdot E\left\{(W_i - EW_i) - \frac{(1-p_i)^n}{r}\right\}\left\{(W_j - EW_j) - \frac{(1-p_j)^n}{r}\right\}$$

$$+ [(1-p_i)^n - (1-p_i-p_j)^n]$$

$$\cdot E\left\{(W_i - EW_i) + \frac{1-(1-p_i)^n}{r}\right\}\left\{(W_j - EW_j) - \frac{(1-p_j)^n}{r}\right\}$$

$$+ [(1-p_j)^n - (1-p_i-p_j)^n]$$

$$\cdot E\left\{(W_i - EW_i) - \frac{(1-p_i)^n}{r}\right\}\left\{(W_j - EW_j) + \frac{1-(1-p_j)^n}{r}\right\}$$

$$+ (1 - p_i - p_j)^n E\left\{(W_i - EW_i) + \frac{1 - (1 - p_i)^n}{r}\right\}$$
$$\cdot \left\{(W_j - EW_j) + \frac{1 - (1 - p_j)^n}{r}\right\}$$
$$= \operatorname{cov} W_i, W_j + \frac{1}{r^2}\{(1 - p_i - p_j)^n - (1 - p_i)^n(1 - p_j)^n\}.$$

Hence

$$\operatorname{var} Z_r^* = \operatorname{var} Z_r + \frac{C^2}{4r^2}\left\{\sum_{i=1}^{k}[(1 - p_i)^n - (1 - p_i)^{2n}] + \sum_{i \neq j=1}^{k}[(1 - p_i - p_j)^n - (1 - p_i)^n(1 - p_j)^n]\right\}$$
$$= \operatorname{var} Z_r + \frac{C^2}{4r^2}\left\{\sum_{i=1}^{k}(1 - p_i)^n - \left[\sum_{i=1}^{k}(1 - p_i)^n\right]^2 + \sum_{i \neq j=1}^{k}(1 - p_i - p_j)^n\right\}.$$

When only one value of Y is observed, which happens with probability $\sum_{i=1}^{k} p_i^n$, the value of Z_r^* is $(C/2)\{(-1)^r/r - 2\sum_{\alpha=r}^{\infty}(-1)^\alpha/\alpha\}$. Since $u_m > 0$, $u_m \to 0$, $u_m > u_{m+1}$ and $u_{m+1} - u_m > u_{m+2} - u_{m+1}$ together imply convergence of $\sum_{m=1}^{\infty}(-1)^{m-1}u_m$ to a value $> u_1/2$, this value of Z_r^* has the sign of $(-1)^{r-1}$. In the case r even, this negative value should be replaced by 0; bias and variance of the resulting modification of Z_r^* are easily found. This point does not arise in Section 3 because if $k = 1$ is known, $H = 0$ is known and estimation is not needed.

A similar discussion can be given for the estimates $H_r''^* = H_r' + C(k^* - 1)/2r$. The maximum likelihood estimates H_r' do not require knowledge of k so can be used unchanged in the case k unknown.

REFERENCES

[1] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1949.

[2] G. A. Miller and W. G. Madow, "On the maximum likelihood estimate of the Shannon-Wiener measure of information," Air Force Cambridge Research Center, 1954.

[3] E. L. Lehmann and H. Scheffé, "Completeness, similar regions and unbiased estimation I," *Sankhya*, Vol. 10 (1950), pp. 305–340.

[4] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley & Sons, 1954.

UNBIASED SEQUENTIAL ESTIMATION FOR BINOMIAL POPULATIONS[1]

By Morris H. DeGroot

Carnegie Institute of Technology

1. Introduction and summary. The subject of minimum variance unbiased estimation has received a great deal of attention in the statistical literature, e.g., in the papers of Bahadur [2], Barankin [3], and Stein [14]. The emphasis in these papers has typically been placed on the existence and construction of minimum variance unbiased estimators when the sampling plan to be used was given in advance.

In this paper, criteria are developed for the selection of an appropriate sampling plan for the family of binomial distributions. Thus, independent observations are to be taken on the random variable U so distributed that

$$\Pr\{U = 1 \mid p\} = p, \qquad \Pr\{U = 0 \mid p\} = 1 - p = q, \tag{1.1}$$

where p lies in the open interval $0 < p < 1$, and the value of a given function, $g(p)$, is to be estimated. The problem considered here is that of determining a sampling plan and an unbiased estimator of $g(p)$ that are optimal, in some sense, at a specified value, p_0, of p. Optimality will depend, not only on the variance of the estimator, but also on the average sample size of the plan. A sampling plan, S, and estimator, f, will be considered optimal at p_0 if, among all procedures with average sample size at p_0 no larger than that of S, there does not exist an unbiased estimator with smaller variance at p_0 than that of f.

The basic tool to be used is the information inequality (see Lemma 2.7 and the discussion following it) which provides a lower bound for the variance of an estimator in terms of its expected value and the average sample size of the sampling plan. If, at p_0, this lower bound is attained for a particular estimator and sampling plan, it may be immediately concluded that they are optimal at p_0. Such an estimator is said to be efficient at p_0.

In Section 2, various definitions, assumptions, and fundamental facts to be used throughout the paper are collected.

In Section 3 it is shown that the single sample plans and the inverse binomial sampling plans are the only ones that admit an estimator that is efficient at all values of p.

In Section 4 some techniques are given that are often useful in the analysis of inverse binomial sampling plans.

In Section 5 relationships between the average sample size of a sampling plan and the functions of p that are estimable optimally are explored.

Received May 26, 1958; revised September 8, 1958.

[1] This paper is based on a dissertation submitted to the Department of Statistics, University of Chicago, in partial fulfillment of the requirements for the Ph.D. degree. The research was sponsored in part by the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

In Sections 6 and 7 it is shown that there can exist two distinct sampling plans with the same average sample size for all p and some comparisons are made of such plans.

In Section 8 a new characterization of completeness is given for bounded sampling plans and it is shown that the dimension of the linear space of unbiased estimators of 0 can be determined simply by counting the number of boundary points. It is further shown that for a wide class of plans, the estimators that are efficient at a given value of p do not have uniformly minimum variance, although non-trivial uniformly minimum variance estimators do exist.

After this paper had been accepted for publication I learned that R. B. Dawson had obtained expression (4.13) and Theorems 8.2 and 8.4, as well as various other interesting results related to the material of Section 8, in his Ph.D. thesis, "Unbiased tests, unbiased estimators, and randomized similar regions," Harvard University, May, 1953.

2. Definitions, assumptions, and fundamental facts. A formal definition of the sampling plans to be considered in the paper will now be given. It will be helpful to keep in mind the interpretation of a sampling plan as a rule that specifies at each stage of a sequential sampling process whether sampling is to cease or another observation is to be taken. Furthermore, it will be helpful to visualize a sequential sample as a path in the Euclidean plane—the path starting at the origin and being extended at a given stage one unit in either the horizontal or vertical direction according as the observation at that stage is 0 or 1. Further discussion of these interpretations will be given below. A formal description of binomial sampling plans was first given by Girshick, Mosteller, and Savage in [7], and the following discussion utilizes several of the concepts presented in that paper and the paper of Lehmann and Stein [11]. In what follows, the word *point* refers only to points γ of the Euclidean plane whose coordinates $X(\gamma)$ and $Y(\gamma)$ are non-negative integers.

A *sampling plan* is a function S defined on the points γ, taking only the values 0 and 1, and such that for the point θ with coordinates $X(\theta) = 0$ and $Y(\theta) = 0$, $S(\theta) = 1$.

A *path* to γ is a sequence of points $\theta = \gamma_0, \gamma_1, \cdots, \gamma_n = \gamma$ such that $S(\gamma_k) = 1$ for $k = 0, 1, \cdots, n-1$ and either

$$\begin{aligned} X(\gamma_{k+1}) &= X(\gamma_k) + 1 \\ Y(\gamma_{k+1}) &= Y(\gamma_k), \end{aligned} \tag{2.1}$$

or

$$\begin{aligned} X(\gamma_{k+1}) &= X(\gamma_k) \\ Y(\gamma_{k+1}) &= Y(\gamma_k) + 1. \end{aligned} \tag{2.2}$$

Thus, each point of the sequence is either one unit to the right or one unit above its predecessor.

Under a given sampling plan, γ is a *boundary point* if there exists a path to γ and $S(\gamma) = 0$. It is a *continuation point* if there exists a path to γ and $S(\gamma) = 1$, so that there is also at least one path "through" γ. It should be noted that the origin θ is always a continuation point.

A point γ is an *inaccessible point* if there does not exist any path to γ. Thus, every point can be uniquely classified as a continuation, inaccessible, or boundary point.

The *sample size* $N(\gamma)$ of any point γ is the sum of its coordinates, $X(\gamma) + Y(\gamma)$.

The *boundary* B is the set of all boundary points.

Clearly, the values of S at inaccessible points are irrelevant to the sampling process. However, some discussions to be given later will be simplified if it is assumed once and for all that $S(\gamma) = 1$ at all inaccessible points. It should be clear that under these conditions S is completely determined by the set of boundary points.

The sampling plans just discussed are somewhat restricted in two respects. The first is that for a given sequence of observations $U_1, \cdots, U_m$ the decision as to whether another observation is taken depends only on the point γ reached by the sample path prescribed by $U_1, \cdots, U_m$; i.e., the decision depends only on m and $\sum_{j=1}^{m} U_j$. The justification for considering only such plans is provided by the fact that the sequence $\sum_{j=1}^{m} U_j$, $m = 1, 2, \cdots$, is both sufficient and transitive. A thorough discussion of these concepts is given by Bahadur in [1]. Similar considerations, especially the theorem of Blackwell [4], justify the definition of an estimator, to be given below, as a statistic that depends on the observed sample sequence only through the boundary point reached by the sequence.

The second restriction on the sampling plans is that they are non-randomized. A randomized sampling plan is a function S defined on the points γ such that S may take any value in the closed unit interval. If, at a given stage in the sampling process, the point γ is reached, then another observation is taken with probability $S(\gamma)$ and sampling is terminated with probability $1 - S(\gamma)$. In this paper, only plans for which $S(\gamma)$ is 0 or 1 are considered.

The probability of reaching a particular boundary point γ is $K(\gamma)p^{Y(\gamma)}q^{X(\gamma)}$, where $K(\gamma)$ is the number of distinct paths to γ.

A sampling plan is said to be *closed* if

$$\sum_{\gamma \in B} K(\gamma)p^{Y(\gamma)}q^{X(\gamma)} = 1 \tag{2.3}$$

for all p, $0 < p < 1$. Only closed plans are considered.

An *estimator* f is a real-valued function defined on B. The only estimators to be considered are those for which

$$E(f \mid p) = \sum_{\gamma \in B} f(\gamma)K(\gamma)p^{Y(\gamma)}q^{X(\gamma)} \tag{2.4}$$

is absolutely convergent.

A sampling plan is *complete* (*boundedly complete*) if the only estimator (bounded

estimator) f, such that $E(f \mid p) = 0$ for all p, is the one defined by $f(\gamma) = 0$ for all $\gamma \in B$.

A sampling plan is *simple*, if, for each positive integer m, the continuation points of sample size m form an interval on the line $X(\gamma) + Y(\gamma) = m$.

The basic facts concerning completeness follow. Lemma 2.1 was developed by Girshick, Mosteller, Savage, and Wolfowitz in a sequence of papers [7], [17], [12]. Lemma 2.2 is due to Lehmann and Stein [11].

LEMMA 2.1. *A necessary and sufficient condition for a closed sampling plan to be boundedly complete is that it be simple.*

LEMMA 2.2. *A necessary and sufficient condition for a closed plan to be complete is that it be simple and that the conversion of any boundary point to a continuation point destroy closure.*

The functions defined on B and taking the values $X(\gamma)$, $Y(\gamma)$, and $N(\gamma)$ are denoted by X, Y, and N, respectively.

The following conditions are assumed to hold throughout the paper:

(i) For every sampling plan to be considered,

$$E(N^2 \mid p) = \sum_{\gamma \in B} N^2(\gamma) K(\gamma) p^{Y(\gamma)} q^{X(\gamma)} \tag{2.5}$$

is uniformly convergent on every closed interval of values of p;

(ii) For every estimator f to be considered,

$$g(p) = E(f \mid p) = \sum_{\gamma \in B} f(\gamma) K(\gamma) p^{Y(\gamma)} q^{X(\gamma)} \tag{2.6}$$

is differentiable termwise in the open interval, $0 < p < 1$, and the derived series is absolutely convergent.

A well-known, elementary, and useful sufficient condition for the termwise differentiability of the series in (2.6) is that the formal termwise derivative be absolutely uniformly convergent on every closed subinterval. This condition is in turn very often verified by a dominance argument (see, e.g., [5], pp. 392, 396).

Some easy but important consequences of the above assumptions will now be given. Some of the results are well-known and have been given elsewhere.

LEMMA 2.3. *$E(Y^2 \mid p)$, $E(X^2 \mid p)$, and $E(XY \mid p)$ all exist and are at most $E(N^2 \mid p)$.*

PROOF. Since $0 \leqq X \leqq N$ and $0 \leqq Y \leqq N$, the results follow from (i).

LEMMA 2.4. *N, X, and Y, considered as estimators, satisfy* (ii).

PROOF. $E(N \mid p) = \sum_{\gamma \in B} N(\gamma) K(\gamma) p^{Y(\gamma)} q^{X(\gamma)}$. The formal termwise derivative of this series is less, in absolute value, than

$$\frac{1}{pq} \sum_{\gamma \in B} N^2(\gamma) K(\gamma) p^{Y(\gamma)} q^{X(\gamma)}.$$

It follows from (i) that this series converges uniformly on every closed interval and, hence, according to the remarks following (ii), $E(N \mid p)$ is termwise differentiable. The proofs for X and Y are similar.

LEMMA 2.5. *If f satisfies* (ii) *and $E(f \mid p) = g(p)$, then $E[(qY - pX)f \mid p] = pqg'(p)$.*

PROOF. Termwise differentiation of the series for $g(p)$ yields the result.

LEMMA 2.6. $qE(Y \mid p) = pE(X \mid p) = E[(qY - pX)^2 \mid p] = pqE(N \mid p)$.

PROOF. These results can be derived directly from Lemma 2.5. However, they are well-known and are basic to various aspects of sequential analysis. As a result, they have been derived under various conditions; e.g., by Wald in the Appendix of [16] as consequences of the fundamental identity of sequential analysis, and by Wolfowitz in [18].

LEMMA 2.7. *For any estimator f,*

$$\text{Var } (f|p) \geqq \frac{pq[g'(p)]^2}{E(N|p)}. \tag{2.7}$$

Equality holds at a particular value of p, say p_0, if and only if there exist constants a and b such that $f(\gamma) = a[q_0 Y(\gamma) - p_0 X(\gamma)] + b$ for all boundary points γ, where $q_0 = 1 - p_0$.

This lemma is also well-known, especially for single sample plans. A brief history of this inequality with references is given by Savage in [13], p. 238. It was first proved for sequential plans by Wolfowitz [18]. Following Savage, (2.7) will be called the *information inequality*.

A non-constant estimator f is said to be *efficient at p_0* if equality holds in the information inequality when $p = p_0$. The expected value of f is then *estimable efficiently at p_0*.

If, for a given sampling plan S, a non-constant estimator f is efficient at p for all p, then both f and S are called *efficient*. From Lemma 2.7, it is seen that f is efficient if and only if there exist two functions of p, $a(p)$ and $b(p)$, with $a(p) \neq 0$, such that

$$f(\gamma) = a(p)[qY(\gamma) - pX(\gamma)] + b(p) \tag{2.8}$$

for all p and all boundary points γ.

Two types of sampling plans of prime importance are the single sample plans and the inverse binomial sampling plans. In a single sample plan,

$$B = \{\gamma : N(\gamma) = n\}$$

for some positive integer n. It is clear that such a plan satisfies (i) since the series involved contains only a finite number of non-zero terms.

In an inverse binomial sampling plan, either $B = \{\gamma : Y(\gamma) = c\}$ for some positive integer c, or $B = \{\gamma : X(\gamma) = c\}$. This type of plan was first treated formally by Haldane in [8] and [9]; the name "inverse binomial sampling" was suggested by Tweedie [15]. It is easily seen that these plans are closed. The next lemma shows that assumption (i) is also satisfied.

LEMMA 2.8. *For an inverse binomial sampling plan $E(N^2 \mid p)$ converges uniformly on every closed interval.*

PROOF. Consider the plan for which $B = \{\gamma : Y(\gamma) = c\}$ for a given positive

integer c. Then

$$\Pr\{N = c + j \mid p\} = \binom{c+j-1}{j} p^c q^j, \qquad j = 0, 1, \cdots,$$

and

$$E(N^2 \mid p) = \sum_{j=0}^{\infty} (c+j)^2 \binom{c+j-1}{j} p^c q^j.$$

On the closed interval $0 < \delta \leqq p \leqq \epsilon < 1$,

$$E(N^2 \mid p) \leqq \epsilon^c \sum_{j=0}^{\infty} (c+j)^2 \binom{c+j-1}{j} (1-\delta)^j,$$

and the series on the right is convergent, as is readily checked by the ratio test. Hence, the given series converges uniformly on the closed interval. The proof is entirely analogous when $B = \{\gamma : X(\gamma) = c\}$.

Because of the duality between the two types of inverse binomial sampling plans, any facts stated in the paper concerning these plans will be demonstrated only for the plan where $B = \{\gamma : Y(\gamma) = c\}$. The analogous proofs for the plan in which $B = \{\gamma : X(\gamma) = c\}$ can always be obtained sim ly by interchanging the roles of X and Y and of p and q.

3. Efficient sampling plans. In this section it is shown that the only efficient sampling plans are the single sample plans and the inverse binomial sampling plans. For a single sample plan, the efficient estimators are the non-constant linear functions of Y, and hence, the functions of p that are estimable efficiently are linear in p. For an inverse binomial sampling plan the efficient estimators are linear functions of N, and their expectations are linear in either $1/p$ or $1/q$.

LEMMA 3.1. *Let S be a given sampling plan for which there exists a non-constant estimator f that is efficient at two values of p. Then there exist constants μ, ν, and ξ, not all 0, such that $\mu X(\gamma) + \nu Y(\gamma) = \xi$ for all boundary points γ.*

PROOF. Suppose f is efficient at p_0 and p_1. Then, from Lemma 2.7, there exist constants a_0, b_0, a_1, and b_1 such that

$$\begin{aligned} f(\gamma) &= a_0[q_0 Y(\gamma) - p_0 X(\gamma)] + b_0 \\ &= a_1[q_1 Y(\gamma) - p_1 X(\gamma)] + b_1 \end{aligned} \tag{3.1}$$

for all boundary points γ. Hence

$$(a_0 q_0 - a_1 q_1) Y(\gamma) - (a_0 p_0 - a_1 p_1) X(\gamma) = b_1 - b_0, \tag{3.2}$$

and since f is not constant, neither a_0 nor a_1 is 0. Thus, not both the coefficients of $Y(\gamma)$ and $X(\gamma)$ are 0, and (3.2) is an equation of the required form.

Clearly, the boundary points of the single sample plans and the inverse binomial sampling plans lie on straight lines as the lemma demands. The following theorem shows that they are the only plans for which this is true.

THEOREM 3.1. *Let S be a given closed sampling plan for which there exist con-*

stants μ, ν, and ξ, not all 0, such that $\mu X(\gamma) + \nu Y(\gamma) = \xi$ for all boundary points γ. Then S is either a single sample plan or an inverse binomial sampling plan.

PROOF. There is no loss of generality in assuming $\mu \geqq 0$. The proof is divided into cases depending on the magnitudes of μ, ν, and ξ.

(i) $\mu = \nu$: If ξ/μ is not a positive integer, then there are no boundary points, and S is not closed. If $\xi/\mu = n$, a positive integer, then B is a subset of the single sample plan with boundary $B^* = \{\gamma : X(\gamma) + Y(\gamma) = n\}$. Clearly, if S is closed, then $B = B^*$.

(ii) $\mu = 0$: If ξ/ν is not a positive integer, then again B is empty, and S is not closed. If $\xi/\nu = c$, a positive integer, then B is a subset of the inverse binomial sampling plan with boundary $B^* = \{\gamma : Y(\gamma) = c\}$. Again, if S is closed, then $B = B^*$.

(iii) $\nu = 0$: As in (ii), B must be of the form $B = \{\gamma : X(\gamma) = c\}$.

(iv) $\mu > 0$, $\nu > 0$, $\mu \neq \nu$: If $\xi \leq 0$, B is empty. If $\xi > 0$, there exists a point γ such that $\mu X(\gamma) + \nu Y(\gamma) < \xi$ and either $\mu[X(\gamma) + 1] + \nu Y(\gamma) > \xi$ or $\mu X(\gamma) + \nu[Y(\gamma) + 1] > \xi$. Thus, with positive probability, the boundary is passed and S is not closed.

(v) $\mu > 0$, $\nu < 0$: Suppose $\xi > 0$. By the strong law of large numbers ([6], p. 243), there is positive probability that, for $\epsilon > 0$,

$$qY(\gamma) - pX(\gamma) > -\epsilon N(\gamma) \tag{3.3}$$

for every point γ reached by the sample path. But for p sufficiently large and ϵ sufficiently small the line $qY(\gamma) - pX(\gamma) = -\epsilon N(\gamma)$ lies entirely above the line $\mu X(\gamma) + \nu Y(\gamma) = \xi$. Hence, no point satisfying the inequality (3.3) can be a boundary point, and there is positive probability that the boundary will not be reached. Thus, S is not closed. Analogous arguments hold if $\xi < 0$ or $\xi = 0$.

The main result of this section is

THEOREM 3.2. *The only efficient sampling plans are the single sample plans and the inverse binomial sampling plans. When $B = \{\gamma : N(\gamma) = n\}$, any non-constant function of the form $a + bY$ is an efficient estimator of $a + bnp$, and these are the only efficient estimators. When $B = \{\gamma : Y(\gamma) = c\}$, any non-constant function of the form $a + bN$ is an efficient estimator of $a + bc(1/p)$, and these are the only efficient estimators. When $B = \{\gamma : X(\gamma) = c\}$, any non-constant function of the form $a + bN$ is an efficient estimator of $a + bc(1/q)$, and these are the only efficient estimators.*

PROOF. Lemma 3.1 and Theorem 3.1 show that the only procedures satisfying certain conditions necessary for a sampling plan to be efficient are the single sample plans and the inverse binomial sampling plans. It remains to show that these procedures are indeed efficient.

Let $B = \{\gamma : N(\gamma) = n\}$. Then, for every $\gamma \in B$, $qY(\gamma) - pX(\gamma) = Y(\gamma) - np$, and hence, $Y(\gamma) = [qY(\gamma) - pX(\gamma)] + np$. This demonstrates the efficiency of Y. Since $E(qY - pX \mid p) = 0$, $E(Y \mid p) = np$.

Let $B = \{\gamma : Y(\gamma) = c\}$. Then, for every $\gamma \in B$, $qY(\gamma) - pX(\gamma) = c - pN(\gamma)$, and hence, $N(\gamma) = -(1/p)[qY(\gamma) - pX(\gamma)] + (c/p)$. Thus, N is efficient and $E(N \mid p) = c/p$.

The proof is completed by noting that if f is an efficient estimator, then so is every non-constant linear function of f. Furthermore, if two non-constant estimators are both efficient at a given value of p, then they are linearly related.

COROLLARY 3.1. *A non-constant estimator is efficient if and only if it is efficient at two distinct values of* p.

PROOF. From Lemma 3.1 and Theorems 3.1 and 3.2, it is seen that if f is not constant and is efficient at two values of p then the sampling plan admits an efficient estimator. From the comments at the end of the proof of Theorem 3.2, it follows that this estimator is linearly related to f, and hence, that f itself is efficient. The proof in the other direction is trivial.

4. Inverse binomial sampling plans. Because of Theorem 3.2, it seems worthwhile to investigate the properties of single sample and inverse binomial sampling plans in some detail.

For the single sample plan with boundary $B = \{\gamma : N(\gamma) = n\}$, Y/n is an efficient estimator of p with $\mathrm{Var}(Y/n \mid p) = pq/n$. For any estimator f,

$$E(f \mid p) = \sum_{\gamma \in B} f(\gamma) \binom{n}{Y(\gamma)} p^{Y(\gamma)} q^{X(\gamma)}, \tag{4.1}$$

which is a polynomial in p of degree at most n. Thus, only polynomials of degree at most n are estimable unbiasedly, and since

$$E\left[\frac{Y(Y-1)\cdots(Y-k)}{n(n-1)\cdots(n-k)} \,\middle|\, p\right] = p^{k+1} \tag{4.2}$$

for $k = 0, 1, \cdots, n-1$, then every such polynomial is estimable.

The analogous properties of an inverse binomial sampling plan are less familiar and will be discussed here.

Consider the plan with boundary $B = \{\gamma : Y(\gamma) = c\}$. For each non-negative integer, k, there exists a unique point, γ_k, of B such that $N(\gamma_k) = c + k$.

Since

$$\Pr\{N = c + k \mid p\} = \binom{k+c-1}{k} p^c q^k, \qquad k = 0, 1, 2, \cdots, \tag{4.3}$$

then for any estimator f,

$$E(f \mid p) = p^c \sum_{k=0}^{\infty} f(\gamma_k) \binom{k+c-1}{k} q^k. \tag{4.4}$$

The class of functions that are estimable unbiasedly and their estimators will now be determined. For convenience, the functions are written as functions of q rather than of p.

THEOREM 4.1. *A function* $h(q)$ *is estimable unbiasedly if and only if it can be expanded in Taylor's series in the interval* $|q| < 1$. *If* $h(q)$ *is estimable, then its unique estimator is given by*

$$f(\gamma_k) = \frac{(c-1)!}{(k+c-1)!} \frac{d^k}{dq^k}\left[\frac{h(q)}{(1-q)^c}\right]_{q=0}, \qquad k = 0, 1, 2, \cdots. \tag{4.5}$$

PROOF. If $h(q)$ can be expanded in Taylor's series in the given interval, then so also can $h(q)/(1-q)^c$, and conversely. Thus, suppose

$$\frac{h(q)}{(1-q)^c} = \sum_{k=0}^{\infty} b_k q^k.$$

Then

$$h(q) = p^c \sum_{k=0}^{\infty} b_k q^k,$$

and taking

$$f(\gamma_k) = b_k \Big/ \binom{k+c-1}{k}$$

yields an estimator f with $E(f \mid p) = h(q)$.

Suppose now that $h(q)$ is estimable unbiasedly. Then there exists an estimator f such that

$$h(q) = p^c \sum_{k=0}^{\infty} f(\gamma_k) \binom{k+c-1}{k} q^k.$$

Replacing p by $1-q$ gives the required expansion. Thus,

$$\frac{h(q)}{(1-q)^c} = \sum_{k=0}^{\infty} f(\gamma_k) \binom{k+c-1}{k} q^k,$$

and hence

$$\binom{k+c-1}{k} f(\gamma_k) = \frac{1}{k!} \frac{d^k}{dq^k} \left[\frac{h(q)}{(1-q)^c} \right]_{q=0},$$

$$f(\gamma_k) = \frac{(c-1)!}{(k+c-1)!} \frac{d^k}{dq^k} \left[\frac{h(q)}{(1-q)^c} \right]_{q=0}.$$

The uniqueness of f follows from the uniqueness of the Taylor's series expansion, which is of course the basis of the completeness of this sampling plan.

Given $h(q)$, Theorem 4.1 provides a rule for finding its unbiased estimator. It is often possible to find the expectation of a given estimator in closed form by using the fact that if the series

$$\phi(x) = \sum_{k=0}^{\infty} b_k x^k \tag{4.6}$$

is differentiated m times within its interval of convergence, then

$$\frac{\phi^{(m)}(x)}{m!} = \sum_{k=0}^{\infty} b_{m+k} \binom{k+m}{k} x^k. \tag{4.7}$$

As illustrations of the technique involved, the variance of the unbiased estimator of p and the moment generating function of N will be determined.

It is well-known (and can be checked from Theorem 4.1) that if f is given by

$$f(\gamma_k) = (c-1)/(k+c-1), \tag{4.8}$$

then, for $c \geqq 2$, f is an unbiased estimator of p. It should be emphasized that this is *not* an efficient estimator. Using the result given above,

$$\begin{aligned} E(f^2 \mid p) &= p^c(c-1)^2 \sum_{k=0}^{\infty} \frac{1}{(k+c-1)^2} \binom{k+c-1}{k} q^k \\ &= p^c(c-1)^2 \frac{1}{(c-1)!} \frac{d^{c-1}}{dq^{c-1}} \left[\sum_{k=1}^{\infty} \frac{1}{k^2} q^k \right]. \end{aligned} \tag{4.9}$$

Note that the constant term in the last series of (4.9) is taken to be 0. Its value can be assigned arbitrarily since it does not appear in the derived series. But

$$\frac{d}{dq}\left[\sum_{k=1}^{\infty} \frac{1}{k^2} q^k\right] = \sum_{k=1}^{\infty} \frac{1}{k} q^{k-1} = \frac{1}{q} \sum_{k=1}^{\infty} \frac{1}{k} q^k = \frac{1}{q}[-\log(1-q)]. \tag{4.10}$$

Hence,

$$\frac{d^{c-1}}{dq^{c-1}}\left[\sum_{k=1}^{\infty} \frac{1}{k^2} q^k\right] = -\frac{d^{c-2}}{dq^{c-2}}\left[\frac{\log(1-q)}{q}\right], \tag{4.11}$$

and

$$E(f^2 \mid p) = -\frac{p^c(c-1)}{(c-2)!} \frac{d^{c-2}}{dq^{c-2}}\left[\frac{\log(1-q)}{q}\right]. \tag{4.12}$$

Using the easily verified fact that

$$\frac{d^m}{dq^m}\left[\frac{\log(1-q)}{q}\right] = \sum_{i=0}^{m} \binom{m}{i} \left[\frac{d^i}{dq^i}(\log(1-q))\right]\left[\frac{d^{m-i}}{dq^{m-i}}\left(\frac{1}{q}\right)\right],$$

one obtains

$$\frac{d^m}{dq^m}\left[\frac{\log(1-q)}{q}\right] = \frac{(-1)^m m! \log(1-q)}{q^{m+1}} + m! \sum_{i=1}^{m} \frac{(-1)^{m-i+1}}{i(1-q)^i q^{m-i+1}}.$$

Thus, referring to (4.12),

$$\begin{aligned} E(f^2 \mid p) &= \frac{(-1)^{c-1}(c-1)p^c \log(1-q)}{q^{c-1}} \\ &\quad + (c-1)p^c \sum_{i=1}^{c-2} \frac{(-1)^{c-i}}{i(1-q)^i q^{c-i-1}} \\ &= \frac{(c-1)p^c}{q^{c-1}}\left[(-1)^{c-1}\log p + \sum_{i=1}^{c-2} \frac{(-1)^{c-i}}{i}\left(\frac{q}{p}\right)^i\right]. \end{aligned} \tag{4.13}$$

It is interesting to note that Haldane [9] gives $E(f^2 \mid p)$ in the form of an integral that, in order to be evaluated, requires repeated integration by parts.

In Section 3 it was shown that N is an efficient estimator of its expected value. Using the technique just illustrated, its moment generating function can be

found. For $t < \log(1/q)$,

$$
\begin{aligned}
E(e^{tN} \mid p) &= p^c \sum_{k=0}^{\infty} e^{t(c+k)} \binom{k+c-1}{k} q^k \\
&= (pe^t)^c \sum_{k=0}^{\infty} \binom{k+c-1}{k} (qe^t)^k \\
&= (pe^t)^c \frac{1}{(c-1)!} \frac{d^{c-1}}{dx^{c-1}} \left[\sum_{k=0}^{\infty} x^k \right]_{x=qe^t} \\
&= (pe^t)^c \frac{1}{(c-1)!} \frac{d^{c-1}}{dx^{c-1}} \left(\frac{1}{1-x} \right)_{x=qe^t} \\
&= (pe^t)^c \frac{1}{(c-1)!} (c-1)! (1-qe^t)^{-c} \\
&= \left(\frac{pe^t}{1-qe^t} \right)^c .
\end{aligned} \tag{4.14}
$$

It should be noted that $N - c$ has a negative binomial distribution and that (4.14) could also be derived using techniques appropriate to that distribution (see, *e.g.*, Feller [6], pp. 155, 252).

Differentiating (4.14) at $t = 0$ yields

$$
\begin{aligned}
E(N \mid p) &= c/p \\
E(N^2 \mid p) &= (c^2 + cq)/p^2 \\
\operatorname{Var}(N \mid p) &= cq/p^2 .
\end{aligned} \tag{4.15}
$$

Since N is an efficient estimator of c/p, its variance should attain the lower bound of the information inequality; *i.e.*,

$$
\operatorname{Var}(N \mid p) = \frac{pq[g'(p)]^2}{E(N \mid p)} . \tag{4.16}
$$

Setting $g(p) = E(N \mid p) = c/p$ does yield the value found in (4.15).

5. Relationships between sampling plans and estimable functions. For a given sampling plan, the only estimators that are efficient at a given value p_0 are the non-constant functions f^* of the form

$$
f^*(\gamma) = a[q_0 Y(\gamma) - p_0 X(\gamma)] + b, \tag{5.1}
$$

for some constants a and b and all boundary points γ.

LEMMA 5.1. *If f^* is given by* (5.1), *then f^* is constant if and only if $a = 0$.*

PROOF. If $a = 0$, then $f^*(\gamma) = b$ for all boundary points γ. If $a \neq 0$, then f^* is constant only if $q_0 Y(\gamma) - p_0 X(\gamma)$ is constant for all boundary points γ. But that would mean that the points of B lie on a straight line and, hence, the sampling plan must be either a single sample plan or an inverse binomial sampling plan. It is readily checked that for neither of these is $q_0 Y(\gamma) - p_0 X(\gamma)$ constant when $0 < p_0 < 1$.

Thus, a non-constant estimator f^* is efficient at p_0 if and only if it is of the form (5.1) with $a \neq 0$. The next theorem determines the class of functions that are estimable efficiently at a given point, simply by evaluating $E(f^* \mid p)$.

THEOREM 5.1. *For a given sampling plan, a non-constant function $g(p)$ is estimable efficiently at p_0 if and only if there exist constants a and b with $a \neq 0$, such that*

$$g(p) = a(p - p_0)E(N \mid p) + b; \tag{5.2}$$

or alternatively, if and only if there exists a constant k, $k \neq 0$, such that

$$\begin{aligned} E(N \mid p) &= k[g(p) - g(p_0)]/(p - p_0) \qquad \text{for } p \neq p_0, \\ E(N \mid p_0) &= kg'(p_0). \end{aligned} \tag{5.3}$$

PROOF. By Lemma 2.6,

$$\begin{aligned} E(q_0Y - p_0X \mid p) &= E(qY - pX + pN - p_0N \mid p) \\ &= E(qY - pX \mid p) + (p - p_0)E(N \mid p) \\ &= (p - p_0)E(N \mid p). \end{aligned}$$

If $g(p)$ is estimable efficiently at p_0 then it must be constant or the expectation of an estimator of the form (5.1) with $a \neq 0$. Hence,

$$g(p) = E(f^* \mid p) = a(p - p_0)E(N \mid p) + b. \tag{5.4}$$

Setting $p = p_0$ in (5.4) gives $b = g(p_0)$. Differentiating both sides of (5.4) at p_0 gives $g'(p_0) = aE(N \mid p_0)$. Thus, both (5.2) and (5.3) are satisfied (with $k = 1/a$).

Conversely, if either (5.2) or (5.3) is satisfied then $g(p)$ is the expectation of an estimator of the form (5.1) with $a = 1/k$ and $b = g(p_0)$.

From Theorem 5.1, it is clear that there does not always exist a sampling plan that admits estimation of a given function efficiently at a given value of p. Since $E(N \mid p)$ can never be smaller than unity, obvious restrictions on $g(p)$ and p_0 are that $[g(p) - g(p_0)]/(p - p_0)$ cannot change sign in the open interval $0 < p < 1$ and that $g'(p_0)$ cannot vanish. However, no general result characterizing the class of functions $\phi(p)$ for which there exist sampling plans with $E(N \mid p) = \phi(p)$ has, to the author's knowledge, yet been found.

6. Some sampling plans. Theorem 5.1 reveals that the functions $g(p)$ that are estimable at a given value p_0 depend only on $E(N \mid p)$ in a very simple way. From the specific form of this relationship it is seen that if it is desired to estimate a polynomial in p efficiently at a particular value, then $E(N \mid p)$ must itself be a polynomial in p of a certain form. In this section some sampling plans for which $E(N \mid p)$ is a polynomial will be described. It should be noted that if, for a given sampling plan, there exists a positive integer n such that $\Pr\{N \leq n \mid p\} = 1$, then $E(N \mid p)$ is a polynomial of degree at most $n - 1$. One of the interesting things about the plans given below is that they yield polynomials for $E(N \mid p)$ even though they do not satisfy this condition. In particular, one of the procedures will be an unbounded sequential plan for which $E(N \mid p)$ is constant.

Hence, considering the single sample plan with the same $E(N \mid p)$, it may be concluded that $E(N \mid p)$ does not determine the sampling plan.

Scheme I. Let n and m be two positive integers and let

$$B = \{\gamma : N(\gamma) = n, Y(\gamma) < n\} \cup \{\gamma : Y(\gamma) = n + m\}. \tag{6.1}$$

That is, a single sample plan of sample size n is used unless all of the observations are equal to 1. When this happens, sampling is continued under an inverse binomial sampling plan until m additional 1's are obtained.

The function $E(N \mid p)$ for the inverse binomial sampling plan was found in Section 4. Using this result, it is easy to obtain $E(N \mid p)$ for the above plan.

$$\begin{aligned} E(N \mid p) &= E(N \mid N \leqq n, p)\Pr(N \leqq n \mid p) + E(N \mid N > n, p)\Pr(N > n \mid p) \\ &= n(1 - p^n) + [n + (m/p)]p^n = n + mp^{n-1}. \end{aligned} \tag{6.2}$$

For example suppose $n = 2$, $m = 4$. Then $E(N \mid p) = 2 + 4p$, and it follows from Theorem 5.1 that p^2 is estimable efficiently at $1/2$. Indeed,

$$\begin{aligned} E(N \mid p) &= 4(p + 1/2) = 4(p^2 - 1/4)/(p - 1/2) \\ &= 4[g(p) - g(p_0)]/(p - p_0), \end{aligned} \tag{6.3}$$

when $g(p) = p^2$, $p_0 = 1/2$. In the same way, $p^2 + (1/2 - p_0)p$ is estimable efficiently at p_0.

If $n = 1$, then, from (6.2), $E(N \mid p) = m + 1$ for all p. From Theorem 5.1 it is seen that this plan admits an efficient estimator of p at any value p_0. This does not mean that the plan is efficient—indeed, it is known not to be. Thus, although for every p_0 there is an estimator of p that is efficient at p_0, there is no single estimator that is efficient at all values of p. As has been shown, such an estimator exists only under a single sample plan. From Lemmas 2.1 and 2.2 it is seen that this plan (and every plan included in Scheme I) is boundedly complete but not complete.

It is interesting to compare the various estimators of p for the plan with $n = 1$. The only bounded unbiased estimator of p is

$$f_0(\gamma) = [1/(m + 1)]Y(\gamma), \tag{6.4}$$

which equals 0 or 1 according as the first observation is 0 or 1. Hence, $\mathrm{Var}(f_0 \mid p) = pq$. Since $E(N \mid p) = m + 1$, $N - (m + 1)$ is an unbounded, unbiased estimator of 0. Hence, any estimator of the form

$$f_c(\gamma) = \frac{1}{m + 1} Y(\gamma) + c\left[1 - \frac{N(\gamma)}{m + 1}\right], \tag{6.5}$$

for some constant c, is an unbiased estimator of p. It is easily checked that f_{p_0} is the efficient estimator at p_0. Thus, the efficient estimators of p are unbounded; this serves as a reminder that unbiased estimators—even efficient ones—are not necessarily desirable estimators.

Using the moments of N found in Section 4 for the inverse binomial sampling

plan, $\mathrm{Var}(f_{p_0} \mid p)$ is found to be

$$\frac{pq}{m+1}\left[1 + m\left(\frac{p_0}{p} - 1\right)^2\right]. \tag{6.6}$$

Comparing this result with the variance of the bounded estimator, f_0, it is seen that $\mathrm{Var}(f_{p_0} \mid p) < \mathrm{Var}(f_0 \mid p)$ for $p > p_0/2$.

Finally, it should be noted that the functions that are estimable efficiently at each p under this plan are precisely the functions that are estimable efficiently under a single sample plan of sample size $m + 1$. The latter plan yields the same $E(N \mid p)$ as the above plan and admits an unbiased estimator of p with uniformly smaller variance than any of the estimators discussed above.

The following variant of Scheme I also yields polynomials for $E(N \mid p)$.

Scheme II. Let n and m be two positive integers with $n \geqq 2$, and let

$$B = \{\gamma : N(\gamma) = n,\ Y(\gamma) < n - 1\} \cup \{\gamma : Y(\gamma) = n + m\}. \tag{6.7}$$

Thus, the boundary is the same as in Scheme I except that the point with coordinates $X(\gamma) = 1$, $Y(\gamma) = n - 1$, is omitted. In other words, a single sample of size n is taken. If all of the observations are equal to 1, sampling is continued until m additional 1's are obtained. If all but one of the initial n observations are equal to 1, sampling is continued until $m + 1$ additional 1's are obtained. If at least two of the initial n observations are not equal to 1, sampling ceases.

Thus, denoting the number of successes in the first n observations by Y_n,

$$\begin{aligned} E(N \mid p) &= E(N \mid Y_n < n - 1, p)\Pr(Y_n < n - 1 \mid p) \\ &\quad + E(N \mid Y_n = n - 1, p)\Pr(Y_n = n - 1 \mid p) \\ &\quad + E(N \mid Y_n = n, p)\Pr(Y_n = n \mid p) \\ &= n(1 - p^n - np^{n-1}q) \\ &\quad + \left(n + \frac{m+1}{p}\right)(np^{n-1}q) + \left(n + \frac{m}{p}\right)p^n \\ &= n + n(m+1)p^{n-2} + (m - n - mn)p^{n-1}. \end{aligned} \tag{6.8}$$

The sampling plans included in Scheme II are also boundedly complete but not complete.

There are obvious variants of Schemes I and II obtained by removing other points on the line $N(\gamma) = n$ from the boundary, or by interchanging the roles of X and Y.

7. Selection of the sampling plan. Theorem 5.1 shows that in order to estimate a function $g(p)$ efficiently at p_0 a sampling plan must be selected for which

$$E(N \mid p) = k[g(p) - g(p_0)]/(p - p_0). \tag{7.1}$$

The efficient estimator at p_0 is defined by

$$f(\gamma) = a[q_0 Y(\gamma) - p_0 X(\gamma)] + g(p_0), \tag{7.2}$$

where $a = 1/k$. As discussed in Section 5, there does not always exist a sampling plan under which a given function $g(p)$ is estimable efficiently at a given value, p_0. On the other hand, it is made clear in Section 6 that for a given $g(p)$ and p_0, there may exist more than one sampling plan satisfying (7.1). Indeed, it has been shown that there may exist more than one plan with a given $E(N \mid p)$ (i.e., with a given value of k).

In this section it is assumed that for a given $g(p)$, p_0, and k, there does exist more than one sampling plan satisfying (7.1). Let $\mathcal{S}$ denote the class of such plans. Since every plan of $\mathcal{S}$ yields the same $E(N \mid p)$, and since, under every plan of $\mathcal{S}$, the estimator f given by (7.2) is efficient at p_0, it follows from the information inequality that $\mathrm{Var}(f \mid p_0)$ is the same under every plan of $\mathcal{S}$. In general, however, for values of p other than p_0, $\mathrm{Var}(f \mid p)$ will be different under the various plans of $\mathcal{S}$. The problem considered here is that of determining the plan of $\mathcal{S}$ for which $\mathrm{Var}(f \mid p)$ is smallest at some value of p other than p_0, or the plan that minimizes $\mathrm{Var}(f \mid p)$ in the neighborhood of p_0. It will be shown that this is equivalent to determining the plan for which $\mathrm{Var}(N \mid p)$ is minimized at the relevant values of p.

In the following derivation of $\mathrm{Var}(f \mid p)$ it is assumed for simplicity that, in (7.2), $a = 1$ and $g(p_0) = 0$. Thus,

$$\begin{aligned} f &= q_0 Y - p_0 X = (qY - pX) + (p - p_0)N, \\ f^2 &= (qY - pX)^2 + (p - p_0)^2 N^2 + 2(p - p_0)N(qY - pX). \end{aligned} \tag{7.3}$$

This yields, upon application of Lemmas 2.5 and 2.6,

$$\begin{aligned} E(f \mid p) &= (p - p_0)E(N \mid p), \\ E(f^2 \mid p) &= pqE(N \mid p) + (p - p_0)^2 E(N^2 \mid p) + 2(p - p_0)pqE'(N \mid p), \end{aligned} \tag{7.4}$$

and hence,

$$\mathrm{Var}(f \mid p) = (p - p_0)^2 \mathrm{Var}(N \mid p) + pqE(N \mid p) + 2(p - p_0)pqE'(N \mid p). \tag{7.5}$$

This expression yields the following theorem, where for any estimator T and sampling plan S, $E(T \mid p, S)$ is the expectation of T at p under the plan S.

THEOREM 7.1. *Let $f = a(q_0 Y - p_0 X) + b$, Let p^* be a value of p other than p_0. Let S_1 and S_2 be two sampling plans such that $E(N \mid p, S_1) = E(N \mid p, S_2)$ for all p. Then* $\mathrm{Var}(f \mid p^*, S_1) \lesseqgtr \mathrm{Var}(f \mid p^*, S_2)$ *if and only if* $\mathrm{Var}(N \mid p^*, S_1) \lesseqgtr \mathrm{Var}(N \mid p^*, S_2)$.

It follows from assumption (i) of Section 2 that $\mathrm{Var}(N \mid p)$ is a continuous function of p. Hence, again referring to (7.5), the following result is immediate.

THEOREM 7.2. *Let f, S_1, and S_2 satisfy the hypotheses of Theorem* 7.1, *and suppose* $\mathrm{Var}(N \mid p_0, S_1) < \mathrm{Var}(N \mid p_0, S_2)$. *Then there exists an interval I of positive length containing p_0 such that*

$$\begin{aligned} &\mathrm{Var}(f \mid p_0, S_1) = \mathrm{Var}(f \mid p_0, S_2), \\ &\mathrm{Var}(f \mid p, S_1) < \mathrm{Var}(f \mid p, S_2), \qquad \text{for } p \in I,\ p \neq p_0. \end{aligned} \tag{7.6}$$

One method that might at first seem useful in determining the plan of $\mathcal{S}$ for which $\mathrm{Var}(N \mid p_0)$ is minimized would be to show that, for a particular plan, this variance attained the lower bound provided by the information inequality; i.e., that

$$\mathrm{Var}\,(N \mid p_0) = \frac{p_0 q_0 [E'(N \mid p_0)]^2}{E(N \mid p_0)}. \tag{7.7}$$

The next lemma and theorem show that there exists such a plan only in the trivial situations where $\mathcal{S}$ contains an efficient sampling plan.

LEMMA 7.1. *For a given sampling plan,*

$$\mathrm{Var}\,(N \mid p) = \frac{pq[E'(N \mid p)]^2}{E(N \mid p)} \tag{7.8}$$

for all p if and only if the sampling plan is efficient.

PROOF. Suppose the plan is efficient. In a single sample plan, N is constant. Hence, $\mathrm{Var}(N \mid p) = E'(N \mid p) = 0$ for all p, and (7.8) holds. In an inverse binomial sampling plan, N is efficient and hence again (7.8) holds.

Suppose the plan is not efficient. Since N is constant only under a single sample plan, it is not constant under the given plan. Thus, (7.8) cannot hold for all p, for if it did, N would be an efficient estimator.

THEOREM 7.3. *For any sampling plan, either*

$$\mathrm{Var}\,(N \mid p) = \frac{pq[E'(N \mid p)]^2}{E(N \mid p)} \qquad \text{for all } p, \tag{7.9}$$

or

$$\mathrm{Var}\,(N \mid p) > \frac{pq[E'(N \mid p)]^2}{E(N \mid p)} \qquad \text{for all } p. \tag{7.10}$$

PROOF. From the information inequality,

$$\mathrm{Var}\,(N \mid p) \geqq \frac{pq[E'(N \mid p)]^2}{E(N \mid p)} \qquad \text{for all } p. \tag{7.11}$$

Suppose equality holds in (7.11) for $p = p_0$. Then N must be of the form

$$N(\gamma) = a[q_0 Y(\gamma) - p_0 X(\gamma)] + b$$

for some constants a and b. Hence,

$$(1 - aq_0)Y(\gamma) + (1 + ap_0)X(\gamma) = b$$

for all boundary points γ. Since the coefficients of X and Y cannot both vanish, the boundary points lie on a straight line, and, by Theorem 3.1, the plan is efficient. The conclusion follows from Lemma 7.1.

It would be reassuring to know that, despite Theorem 7.3, there always exists a plan in $\mathcal{S}$ for which $\mathrm{Var}(N \mid p_0)$ is smallest. This would be true if $\mathcal{S}$ contained only a finite number of sampling plans. No general results have as yet been

obtained in this direction, but the following theorem concerning the specific case where $\mathcal{S}$ contains all plans with $E(N \mid p) = 2$ indicates the type of result that might be expected.

THEOREM 7.4. *There exist exactly three sampling plans for which* $E(N \mid p) = 2$ *for all* p.

PROOF. Three plans have already been given for which $E(N \mid p) = 2$; namely, the single sample plan, the plan S^* given under Scheme I of Section 6, with $n = 1$, $m = 1$, and the symmetric image of this plan obtained by interchanging the roles of X and Y. It will now be shown that these are the only three plans. Throughout the proof, points will be specified by their coordinates $(X(\gamma), Y(\gamma))$.

Consider a sampling plan S with boundary B for which $E(N \mid p) = 2$ for all p. The points $(0, 1)$ and $(1, 0)$ cannot both be in B, for then $E(N \mid p) = 1$. If neither $(0, 1)$ nor $(1, 0)$ is in B, then $N \geqq 2$. Thus, if $E(N \mid p) = 2$, then $\Pr(N = 2 \mid p) = 1$ and S is the single sample plan. Suppose then that exactly one of the points $(0, 1)$ and $(1, 0)$ is in B and, for the moment, assume it to be $(1, 0)$. It will be shown that S is the plan S^* described above with boundary

$$B^* = \{(1, 0), (0, 2), (1, 2), (2, 2), (3, 2), \cdots\}.$$

Suppose $(0, 2) \notin B$. Then $N \geqq 3$ whenever the sample path goes through $(0, 2)$. Hence, $E(N \mid p) \geqq 3p^2$, which is greater than 2 for values of p arbitrarily close to 1. Thus, $(0, 2) \in B$. Also, $(1, 1) \notin B$, for, otherwise, $\Pr\{N \leqq 2 \mid p\} = 1$, $\Pr\{N < 2 \mid p\} > 0$, and $E(N \mid p) < 2$. The proof is now completed by an inductive argument.

Assume, for a given integer n, $n \geqq 0$, the points $(1, 0)$ and $(0, 2), (1, 2), \cdots, (n, 2)$ are in B and the points $(0, 1), (1, 1), \cdots, (n + 1, 1)$ are not in B. It must be shown that $(n + 1, 2) \in B$ and $(n + 2, 1) \notin B$.

Suppose $(n + 1, 2) \notin B$. Then $N \geqq n + 4$ whenever the sample path goes through $(n + 1, 2)$. This happens with probability p^2q^{n+1}. Thus, by considering the other points of B,

$$\begin{aligned} E(N \mid p) &\geqq q + 2p^2 + 3p^2q + \cdots + (n + 2)p^2q^n + (n + 4)p^2q^{n+1} \\ &= q + p^2(1 + 2q + \cdots + (n + 1)q^n + (n + 2)q^{n+1}) \\ &\qquad + p^2(1 + q + \cdots + q^n + q^{n+1}) + p^2q^{n+1} \\ &= q + p^2 \frac{d}{dq}\left[\frac{1 - q^{n+3}}{1 - q}\right] + p^2\left[\frac{1 - q^{n+2}}{1 - q}\right] + p^2q^{n+1} \\ &= 2 + p^2q^{n+1} - (n + 4)pq^{n+2} - q^{n+3}. \end{aligned}$$

Now let $p = 1 - \delta$, $q = \delta$. Then

$$E(N \mid p) \geqq 2 + \delta^{n+1} + o(\delta^{n+1}),$$

and for δ positive but arbitrarily close to 0, $E(N \mid p) > 2$. It follows that $(n + 1, 2) \in B$.

Now, if $(n + 2, 1)$ were in B, then S would be a bounded procedure and its

continuation points would be a proper subset of the continuation points of S^*. Since S and S^* yield the same $E(N \mid p)$, this is impossible. Thus, $(n + 2, 1) \varepsilon B$ and the induction is completed. It follows that $S = S^*$.

If, originally, it was assumed that $(0, 1) \varepsilon B$, $(1, 0) \varepsilon B$, then an entirely analogous demonstration would show that S must be the symmetric image of S^* described at the beginning of the proof.

8. On the completeness of bounded sampling plans. A *bounded* sampling plan is one for which there exists a positive integer n such that

$$\Pr\{N \leqq n \mid p\} = 1. \tag{8.1}$$

The *size* of a bounded sampling plan is the smallest n for which (8.1) holds.

In the following Theorems 8.1–8.5 it is shown that if the boundary of a plan of size n contains $n + 1 + k$ points, $k \geqq 0$, then k is the dimension of the linear space of unbiased estimators of 0.

THEOREM 8.1. *The boundary of a sampling plan of size n contains at least* $n + 1$ *points.*

PROOF. If $n = 1$, then (0, 1) and (1, 0) must both be boundary points. Proceeding by induction, suppose the theorem to be true for $n = m$ and consider a plan S of size $m + 1$. Since $m + 1 \geqq 2$, the points (0, 1) and (1, 0) cannot both be boundary points and there must exist a path through at least one of them that extends to a boundary point of sample size $m + 1$. Without loss of generality, suppose this is true of the point (1, 0). Then, given that the sample path has reached (1, 0), the sampling plan (or what remains of it) is now of size m and hence, by the induction hypothesis, involves at least $m + 1$ boundary points. In other words, there are at least $m + 1$ boundary points that can be reached by paths through (1, 0). Since S is closed and bounded there must also exist a boundary point of the form $(0, y)$. Since this point cannot be reached from (1, 0) it is not counted above, and the boundary contains at least $m + 2$ points.

THEOREM 8.2. *If the boundary of a sampling plan of size n contains more than* $n + 1$ *points, the plan is not complete.*

PROOF. The probability, $P(p; \gamma)$, of reaching a particular boundary point γ is

$$P(p; \gamma) = K(\gamma)p^{r(\gamma)}q^{x(\gamma)},$$

a polynomial in p of degree at most n. Hence, the expectation

$$\sum_{\gamma \varepsilon B} f(\gamma)P(p; \gamma)$$

of any estimator f is also such a polynomial. Conversely, every linear combination of the $P(p; \gamma)$ can be attained as the expectation of some estimator. Thus, there exists a non-trivial estimator f such that $E(f \mid p) = 0$ for all p if and only if there exist constants $a(\gamma)$, not all 0, such that $\sum_{\gamma \varepsilon B} a(\gamma)P(p; \gamma) = 0$ for all p. But the linear space of polynomials in p of degree at most n has dimension $n + 1$ and hence, if B contains more than $n + 1$ points, the set of polynomials $\{P(p; \gamma) : \gamma \varepsilon B\}$ must be linearly dependent.

THEOREM 8.3. *Under a sampling plan of size n, all polynomials in p of degree at most n are estimable unbiasedly.*

PROOF. The polynomial 1 is trivially estimable unbiasedly. Furthermore, since the plan is of size n, there exists a boundary or continuation point γ_j of sample size j for each j, $1 \leqq j \leqq n$. Let $\gamma_j = (x_j, y_j)$, where $x_j + y_j = j$. Girshick Mosteller, and Savage [7], have shown explicitly how to construct an unbiased estimator of $p^{y_j}q^{x_j}$, a polynomial in p of degree j. Thus, the polynomials $1, p^{y_1}q^{x_1}, \cdots, p^{y_n}q^{x_n}$ are all estimable unbiasedly and, hence, so is any linear combination of them. Since no two of these polynomials are of the same degree they are linearly independent and thus form a basis for the $n + 1$-dimensional space of all polynomials in p of degree at most n. It follows that every such polynomial is estimable unbiasedly.

THEOREM 8.4. *If the boundary of a sampling plan of size n contains exactly $n + 1$ points the plan is complete.*

PROOF. The expectation of every estimator is a linear combination of the $n + 1$ polynomials $P(p; \gamma)$, $\gamma \varepsilon B$, and by Theorem 8.3, every polynomial of degree at most n can be expressed as such a linear combination. Thus, the set $\{P(p; \gamma) : \gamma \varepsilon B\}$ spans the $n + 1$-dimensional linear space of polynomials. It follows that the $n + 1$ polynomials $P(p; \gamma)$ must be linearly independent and, hence, there exist no non-trivial unbiased estimators of 0.

THEOREM 8.5. *If the boundary of a sampling plan of size n contains $n + 1 + k$ points, $k > 0$, then there exist exactly k linearly independent unbiased estimators of* 0.

PROOF. Consider the $n + 1 + k$ boundary points $\gamma_1, \cdots, \gamma_{n+1+k}$. Each estimator f can be considered as a vector $(f_1, \cdots, f_{n+1+k})$ where $f(\gamma_j) = f_j$, $j = 1, \cdots, n + 1 + k$. Thus, the space of estimators can be considered as an $n + 1 + k$-dimensional linear space, V. The expectation operator E is a linear mapping from V onto the $n + 1$-dimensional linear space of polynomials in p of degree at most n. The subspace $V^0 = \{f : E(f \mid p) = 0 \text{ for all } p\}$ is the null space of this mapping and it follows from the standard theorems concerning rank and nullity that V^0 has dimension k.

The remainder of this section is devoted to bounded sampling plans that are complete only after the removal of some points from the boundary. It will be shown that for a plan of this type it is easy to explicitly construct a basis for the space of unbiased estimators of 0.

The following notation is used in Theorems 8.6–8.8. It is assumed that S is a bounded sampling plan with boundary B, and that $\beta_1, \cdots, \beta_t$, with $t > 0$, are t points of B such that the sampling plan with boundary $B - \{\beta_1, \cdots, \beta_t\}$ is closed and complete.

The sampling plan with boundary $B_j = B - \{\beta_j\}$ is denoted by S_j for $j = 1, \cdots, t$.

For any point $\gamma \varepsilon B$, $K(\gamma)$ is the number of paths to γ under the plan S and $K_j(\gamma)$ is the number of paths to γ under the plan S_j for $j = 1, \cdots, t$. If γ is not a boundary point for a particular plan then the number of paths to γ under that plan is taken to be 0. Thus, $K_j(\beta_j) = 0$ for $j = 1, \cdots, t$.

It should be noted that $K(\gamma)$ does not vanish for any $\gamma \varepsilon B$. Also, if $\gamma \neq \beta_j$ then $K_j(\gamma) \geqq K(\gamma)$. This follows from the fact that any path to γ under S is obviously a path to γ under any sampling plan with fewer boundary points.

Consider now, under the sampling plan S, the estimators ϕ_j defined for $j = 1, \cdots, t$ by

$$\phi_j(\gamma) = 1 - [K_j(\gamma)/K(\gamma)], \tag{8.2}$$

for $\gamma \varepsilon B$.

THEOREM 8.6. *Under the sampling plan S, the estimators $\phi_1, \cdots, \phi_t$ form a basis for the linear space of unbiased estimators of* 0.

PROOF. It was shown by Girshick, Mosteller, and Savage [7], and it is very easy to verify, that, for each j, ϕ_j is not identically 0 and $E(\phi_j \mid p) = 0$ for all p. Since the boundary of S contains t more points than that of a complete plan of the same size as S it follows from Theorem 8.5 that the dimension of the space of unbiased estimators of 0 is t. Thus, the theorem will be proven if it is shown that the estimators $\phi_1, \cdots, \phi_t$ are linearly independent.

Let the sample size of β_j be n_j (i.e., $X(\beta_j) + Y(\beta_j) = n_j$) and assume that $\beta_1, \cdots, \beta_t$ have been ordered so that $n_1 \leqq n_2 \leqq \cdots \leqq n_t$. Consider the matrix A with elements $a_{ij} = \phi_j(\beta_i)$; thus,

$$A = \begin{pmatrix} \phi_1(\beta_1) & \cdots & \phi_t(\beta_1) \\ \vdots & & \vdots \\ \phi_1(\beta_t) & \cdots & \phi_t(\beta_t) \end{pmatrix}. \tag{8.3}$$

For $j > i$, $K_j(\beta_i) = K(\beta_i)$. This follows by noting that a path to β_i under S_j goes through points of sample size less than n_i and then through β_i. It does not go through β_j since $n_j \geqq n_i$, and hence, it is also a path to β_i under S. Thus, $K_j(\beta_i) \leqq K(\beta_i)$. The reverse inequality has been stated in the comments preceding the theorem. It follows from (8.2) that $\phi_j(\beta_i) = 0$ for $j > i$. Since $\phi_j(\beta_j) = 1$ for each j, the matrix A is seen to be triangular with each of its diagonal elements equal to 1. Hence, $|A| = 1$. It can now be concluded that $\phi_1, \cdots, \phi_t$ are linearly independent, for otherwise $|A|$ must vanish. This completes the proof.

Under a given sampling plan, an estimator f^* is said to be a *uniformly minimum variance estimator* of its expected value, if, for any other estimator f with the same expectation as f^*, $\mathrm{Var}(f^* \mid p) \leqq \mathrm{Var}(f \mid p)$ for all p. Consider the estimator $f^* = q_0 Y - p_0 X$ that is efficient at p_0. The following theorem shows that under a sampling plan of the type now being considered f^* is not a uniformly minimum variance estimator. On the other hand, if f is any other estimator with the same expectation as f^*, then $\mathrm{Var}(f^* \mid p_0) < \mathrm{Var}(f \mid p_0)$ and f is not a uniformly minimum variance estimator. Thus, under the sampling plan S, if a non-constant function $g(p)$ is estimable efficiently at p_0 then there is no uniformly minimum variance estimator of $g(p)$.

THEOREM 8.7. *Under the sampling plan S, the estimator $q_0 Y - p_0 X$ is not a uniformly minimum variance estimator.*

PROOF. Lehmann and Scheffé [10] have shown that for bounded sampling plans a necessary and sufficient condition for an estimator T to be a uniformly minimum variance estimator is that $E(T\phi \mid p) = 0$ for every ϕ such that $E(\phi \mid p) = 0$. Suppose $q_0Y - p_0X$ was a uniformly minimum variance estimator. Then, for $j = 1, \cdots, t$,

$$\begin{aligned} 0 &= E[\phi_j(q_0Y - p_0X) \mid p] \\ &= E[\phi_j(qY - pX) \mid p] + (p - p_0)E(\phi_jN \mid p). \end{aligned} \tag{8.4}$$

By Lemma 2.5, $E[\phi_j(qY - pX) \mid p] = 0$ for all p, and hence, $E(\phi_jN \mid p) = 0$ for all p and each j. (Actually, equation (8.4) yields $E(\phi_jN \mid p) = 0$ only for $p \neq p_0$. That $E(\phi_jN \mid p_0) = 0$ follows from continuity.) In particular, $E(\phi_tN \mid p) = 0$ where again it is assumed that the points $\beta_1, \cdots, \beta_t$ have been ordered so that $n_1 \leqq \cdots \leqq n_t$.

It follows from Theorem 8.6 that there exist constants $r_1, \cdots, r_t$ such that

$$N(\gamma)\phi_t(\gamma) = \sum_{j=1}^{t} r_j\phi_j(\gamma) \tag{8.5}$$

for all $\gamma \varepsilon B$. In particular,

$$n_t\phi_t(\beta_t) = \sum_{j=1}^{t} r_j\phi_j(\beta_t) \tag{8.6}$$

for $i = 1, \cdots, t$. Recalling that $\phi_j(\beta_j) = 1$ and $\phi_j(\beta_i) = 0$ for $j > i$, the values of $r_1, \cdots, r_t$ are readily found from (8.6) to be $r_1 = \cdots = r_{t-1} = 0$, $r_t = n_t$. Thus, from (8.5),

$$N(\gamma)\phi_t(\gamma) = n_t\phi_t(\gamma) \tag{8.7}$$

for all $\gamma \varepsilon B$. It follows that $\phi_t(\gamma) = 0$ for all γ such that $N(\gamma) \neq n_t$. But, as argued in the proof of Theorem 8.6, $\phi_t(\gamma) = 0$ for all boundary points of sample size n_t with the exception of β_t. Hence,

$$E(\phi_t \mid p) = \phi_t(\beta_t)K(\beta_t)p^{Y(\beta_t)}q^{X(\beta_t)} \neq 0, \tag{8.8}$$

a contradiction. It follows that $q_0Y - p_0X$ is not a uniformly minimum variance estimator.

The next theorem shows that, despite Theorem 8.7, there do exist non-constant uniformly minimum variance estimators.

THEOREM 8.8. *Under the sampling plan S, there exist non-constant uniformly minimum variance estimators.*

PROOF. Since S is bounded, there exists a boundary point $\alpha_0 = (x, 0)$ and a boundary point $\alpha_1 = (0, y)$. Neither α_0 nor α_1 is one of the points $\beta_1, \cdots, \beta_t$ since if either is removed from B the resulting plan is not closed. Furthermore, $\phi_j(\alpha_0) = \phi_j(\alpha_1) = 0$ for $j = 1, \cdots, t$ since, under any of the plans S or S_j, there is only one path to either α_0 or α_1. Now consider an estimator f of the form

$$
\begin{aligned}
f(\alpha_0) &= f_0\,, \\
f(\alpha_1) &= f_1\,, \\
f(\gamma) &= c \qquad \text{for } \gamma \in B - \{\alpha_0\,, \alpha_1\}.
\end{aligned} \tag{8.9}
$$

Then $E(f\phi_j \mid p) = cE(\phi_j \mid p) = 0$ for $j = 1, \cdots, t$, and it follows from Theorem 8.6 and the condition of Lehmann and Scheffé given at the beginning of the proof of Theorem 8.7 that f is a uniformly minimum variance estimator.

9. Acknowledgments. I am deeply indebted to L. J. Savage, under whose careful and considerate guidance this research was done, and to S. G. Ghurye and W. H. Kruskal for several helpful comments.

REFERENCES

[1] R. R. Bahadur, "Sufficiency and statistical decision functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 423–462.

[2] R. R. Bahadur, "On unbiased estimates of uniformly minimum variance," *Sankhya*, Vol. 18 (1957), pp. 211–224.

[3] E. W. Barankin, "Locally best unbiased estimates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 477–501.

[4] D. Blackwell, "Conditional expectation and unbiased sequential estimation," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 105–110.

[5] R. Courant, *Differential and Integral Calculus*, Vol. 1, Interscience Publishers, Inc., New York, 1937.

[6] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 2nd Edition, John Wiley and Sons, New York, 1957.

[7] M. A. Girshick, F. Mosteller, and L. J. Savage, "Unbiased estimates for certain binomial sampling problems with applications," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 13–23.

[8] J. B. S. Haldane, "A labour-saving method of sampling," *Nature*, Vol. 155 (1945), pp. 49–50.

[9] J. B. S. Haldane, "On a method of estimating frequencies," *Biometrika*, Vol. 33 (1945), pp. 222–225.

[10] E. L. Lehmann and H. Scheffé, "Completeness, similar regions and unbiased estimation—Part I," *Sankhya*, Vol. 10 (1950), pp. 305–340.

[11] E. L. Lehmann and C. Stein, "Completeness in the sequential case," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 376–385.

[12] L. J. Savage, "A uniqueness theorem for unbiased sequential estimation," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 295–297.

[13] L. J. Savage, *The Foundations of Statistics*, John Wiley and Sons, New York, 1954.

[14] C. Stein, "Unbiased estimates with minimum variance," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 406–415.

[15] M. C. K. Tweedie, "Inverse statistical variates," *Nature*, Vol. 155 (1945), p. 453.

[16] A. Wald, *Sequential Analysis*, John Wiley and Sons, New York, 1947.

[17] J. Wolfowitz, "On sequential binomial estimation," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 489–493.

[18] J. Wolfowitz, "The efficiency of sequential estimates and Wald's equation for sequential processes," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 215–230.

A SINGLE-SAMPLE MULTIPLE-DECISION PROCEDURE FOR SELECTING THE MULTINOMIAL EVENT WHICH HAS THE HIGHEST PROBABILITY[1]

By Robert E. Bechhofer, Salah Elmaghraby, and Norman Morse

Sibley School of Mechanical Engineering, Cornell University

Summary. The problem of selecting the multinomial event which has the highest probability is formulated as a multiple-decision selection problem. Before experimentation starts the experimenter must specify two constants (θ^*, P^*) which are incorporated into the requirement: "The probability of a correct selection is to be equal to or greater than P^* whenever the true (but unknown) ratio of the largest to the second largest of the population probabilities is equal to or greater than θ^*." A single-sample procedure which meets the requirement is proposed. The heart of the procedure is the proper choice of N, the number of trials. Two methods of determining N are described: the first is exact and is to be used when N is small; the second is approximate and is to be used when N is large. Tables and sample calculations are provided.

1. Introduction. We are concerned in this paper with the multiple-decision problem which arises when one attempts to answer questions such as the following:

(a) Which of the six faces of a loaded die has the largest probability of landing face up?

(b) Which of the thirty-six "bettable" numbers on an unbalanced roulette wheel has the largest probability associated with it?

(c) Which of the k television programs available to a given TV audience in a certain locale can claim the largest proportion of the total audience as listeners?

The multinomial distribution provides a statistical model for dealing with each of these questions. In the following sections it is shown how such questions as these can be formulated as *multiple-decision selection problems*. A single-sample procedure is proposed which provides a solution to these problems.

2. Statistical assumptions, and definitions. Let $\mathbf{X}_j = (X_{1j}, X_{2j}, \cdots, X_{kj})$ be independent vector-observations from the same multinomial population with a common unknown probability vector $\mathbf{p} = (p_1, p_2, \cdots, p_k)$; here p_i is the probability of the event $E_i (0 \leqq p_i \leqq 1, \sum_{i=1}^{k} p_i = 1)$ and $X_{ij} = 1$ or 0 according as E_i does or does not occur on the jth observation $(i = 1, 2, \cdots, k; j = 1, 2, \cdots)$. Let $p_{[1]} \leqq p_{[2]} \leqq \cdots \leqq p_{[k]}$ denote the ranked probabilities. It is assumed

Received February 5, 1958; revised March 25, 1958.

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 18(600)-331. Reproduction in whole or in part is permitted for any purpose of the United States Government.

that the experimenter has no a priori knowledge which would rule out any of the $k!$ possible pairings of the $p_{[i]}$ with the $E_i (i = 1, 2, \cdots, k)$. Let $\theta_{i,j} = p_{[i]}/p_{[j]}$ $(i \geqq j; i, j = 1, 2, \cdots, k)$. Let N be the number of vector observations; let $Y_{iN} = \sum_{j=1}^{N} X_{ij} (i = 1, 2, \cdots, k)$, and let $y_{[1]N} \leqq y_{[2]N} \leqq \cdots \leqq y_{[k]N}$ denote their ranked values. Let $E_{[i]N}$ denote the event associated with $y_{[i]N}$. Let $Y_{(i)N}$ be that one of the Y_{iN} which is associated with the event having probability $p_{[i]}$ $(i = 1, 2, \cdots, k)$.

3. Goal, specification, and requirement. We now state the experimenter's goal, specification, and requirement:

Goal. The experimenter's goal is to select the event associated with $p_{[k]}$.

The statistical formulation of the problem for this goal involves the true ratio $\theta_{k,k-1} = \theta$ (say) and the true probability P of a correct selection. It is assumed that *before experimentation starts* the experimenter can specify a pair of constants (θ^*, P^*) with $1 < \theta^* < \infty$ and $1/k < P^* < 1$ as described below.

Specification. The experimenter specifies:

(a) The smallest value, θ^*, of the ratio θ that is worth detecting, and

(b) The smallest acceptable value, P^*, of the probability P of achieving the above goal when $\theta \geqq \theta^*$.

The specification above is summarized in the following:

Requirement. The experimenter requires that the procedure to be used guarantee that

$$\text{Probability } \{\text{Correct selection} \mid \theta_{k,k-1} \geqq \theta^*\} \geqq P^*.$$

That is, the probability of a correct selection is to be equal to or greater than P^* whenever the true (but unknown) ratio of the largest to the second largest of the population probabilities p_i is equal to or greater than θ^*.

4. Procedure. We propose a single-sample procedure that will guarantee the requirement. It is similar to ones described in [1], [2], and [6] which are used to solve selection problems involving parameters associated with other basic distributions. (A *sequential* procedure which will guarantee the same requirement (above) was reported on in [3]; the theory underlying the sequential procedure will be given in a paper with the same title as [3] which is being prepared by the authors of that abstract. Either the sequential procedure reported on in [3] or the single-sample procedure described in the present paper can also be used to solve problems of the type posed in [4].)

Our single-sample procedure takes the following form:

Procedure.

(a) Select a random sample of N vector-observations.

(b) Compute the $y_{[i]N} (i = 1, 2, \cdots, k)$.

(c) If *exactly* s $(s = 1, 2, \cdots, k)$ of the $y_{[i]N}$ are tied for largest, $y_{[k-s+1]N} = y_{[k-s+2]N} = \cdots = y_{[k]N}$, select as the event associated with $p_{[k]}$, one of $E_{[k-s+1]N}$, $E_{[k-s+2]N}$, $\cdots$, $E_{[k]N}$ using a random device which assigns probability $1/s$ to

each of them. In particular, if there is a *single* largest one, *i.e.*, if $s = 1$, select $E_{[k]N}$.

The heart of the problem in terms of designing the experiment is the proper choice of N; it must be chosen just large enough to guarantee the requirement, i.e., P^* must be achieved for all possible $\mathbf{p} = (p_1, p_2, \cdots, p_k)$ for which $\theta_{k,k-1} \geqq \theta^*$. In order to accomplish this we consider the *least favorable configuration* (l.f.c.) of the $p_{[i]}$'s the definition of which follows.

We define the l.f.c. of the $p_{[i]}$'s as that probability vector $\mathbf{p}$ which for any given N, k, θ^* minimizes the probability of a correct selection when $\theta_{k,k-1} \geqq \theta^*$. It is proved in [7] that:

(a) The l.f.c. is independent of N.

(b) The l.f.c. is given by

$$\theta_{k,i} = \theta^* \qquad (i = 1, 2, \cdots, k-1). \tag{1}$$

Since $\sum_{i=1}^{k} p_{[i]} = 1$, this implies

$$\begin{aligned} p_{[1]} &= p_{[2]} = \cdots = p_{[k-1]} = 1/(\theta^* + k - 1) \\ p_{[k]} &= \theta^*/(\theta^* + k - 1). \end{aligned} \tag{2}$$

It is intuitively clear that for any $\mathbf{p}$ for which the p_i are not all equal, the probability of a correct selection increases with N. Hence if N is chosen large enough to guarantee a specified probability of a correct selection when the configuration is least favorable, it will guarantee *at least* that probability for *any* configuration of $p_{[i]}$'s with $\theta_{k,k-1} \geqq \theta^*$, i.e., it will be large enough to guarantee the requirement. We shall denote by N^* the smallest value of N which will guarantee the requirement.

In the next section we shall show how to compute the *exact* probability of a correct selection for any k, N, and $\mathbf{p}$; this method is appropriate when N is small. Methods of approximating these probabilities when N is not too small are given in sections 6 and 7. Using these methods it is possible to compute tables from which N^* can be determined for any specification. Several such tables are given at the end of this paper.

5. Exact probability of a correct selection. For any fixed k and N and any probability vector $\mathbf{p}$ with $p_{[k-1]} < p_{[k]}$, the probability of a correct selection is given by

$$\phi_k = \phi(p_{[1]}, \cdots, p_{[k]}) = \sum \frac{1}{s} \cdot \frac{N!}{y_{(1)N}! \cdots y_{(k)N}!} p_{[1]}^{y_{(1)N}} \cdots p_{[k]}^{y_{(k)N}} \tag{3}$$

where the summation is over all vectors $\mathbf{y}_N = (y_{(1)N}, y_{(2)N}, \cdots, y_{(k)N})$ such that $\sum_{i=1}^{k} y_{(i)N} = N$ and $y_{(k)N} \geqq y_{(i)N}$ $(i = 1, 2, \cdots, k-1)$, and s, which is a function of $\mathbf{y}_N$, is the number of $y_{(i)N}$'s tied for largest. (Note: If *exactly* s $(s = 1, 2, \cdots, k)$ of the $p_{[i]}$ are tied for largest, then $p_{[k-s+1]} = p_{[k-s+2]} = \cdots = p_{[k]}$, and we consider the selection of any one of the associated s events as a correct selection.)

For fixed k and N it is straightforward (but tedious) to express ϕ_k in terms of the $p_{[i]}(i = 1, 2, \cdots, k)$; for example, if $k = 3$, $N = 4$ and $p_{[3]} > p_{[2]}$, it is easy to verify that

$$\phi(p_{[1]}, p_{[2]}, p_{[3]}) = p_{[3]}^4 + 4p_{[3]}^3(1 - p_{[3]}) + 3p_{[3]}^2(p_{[3]}^2 + 4p_{[3]}p_{[1]} + p_{[1]}^2) \tag{4}$$

and if the $p_{[i]}$ $(i = 1, 2, 3)$ are in the l.f.c. (2), this reduces to

$$\phi\left(\frac{1}{\theta^* + 2}, \frac{1}{\theta^* + 2}, \frac{\theta^*}{\theta^* + 2}\right) = \frac{(\theta^*)^2}{(\theta^* + 2)^4}\{(\theta^*)^2 + 8\theta^* + 18\} \tag{5}$$

which involves only θ^*. Substituting the numerical value of the specified θ^* in (5) gives the *exact* probability of a correct selection for $k = 3$ and $N = 4$ when the $p_{[i]}$ $(i = 1, 2, 3)$ are in the l.f.c. (2). General expressions of the type of (4) can be obtained for arbitrary k and N, and using (2) these can be reduced to expressions of the type of (5) which involve only θ^*. The exact probabilities of a correct selection which are listed in Tables A-2, A-3, and A-4 are all associated with the l.f.c. and were computed using expressions of the type of (5). It should be emphasized, however, that it can be extremely laborious to express (3) in the form (5), especially when N is moderately large. For example, for $k = 4$ and $N = 20$ the expression analogous to the term in braces in (5) is a fifteenth-degree polynomial in which five of the coefficients are eleven digit integers. The evaluation of such expressions for the specified θ^* is an additional problem (although this can be done expeditiously on a high-speed electronic computer). It thus is obvious that some large-sample approximation for (3) would be desirable.[2] We consider this problem in the next sections.

6. Large-sample approximations to the probability of a correct selection. First we shall consider the case $k = 3$ to see what might be involved in making a large-sample approximation. We have

$$\begin{aligned} \phi_3 = {} & \Pr\{Y_{(3)N} - Y_{(1)N} > 0, Y_{(3)N} - Y_{(2)N} > 0\} \\ & + \tfrac{1}{2}\Pr\{Y_{(3)N} - Y_{(1)N} = 0, Y_{(3)N} - Y_{(2)N} > 0\} \\ & + \tfrac{1}{2}\Pr\{Y_{(3)N} - Y_{(1)N} > 0, Y_{(3)N} - Y_{(2)N} = 0\} \\ & + \tfrac{1}{3}\Pr\{Y_{(3)N} - Y_{(1)N} = 0, Y_{(3)N} - Y_{(2)N} = 0\}. \end{aligned} \tag{6}$$

Because of the equality signs, the last three terms become negligible for large N; hence, for large N we shall approximate ϕ_3 by

$$\phi_3' = \Pr\{Y_{(3)N} - Y_{(1)N} \geqq 0, Y_{(3)N} - Y_{(2)N} \geqq 0\}. \tag{7}$$

[2] The referee has called our attention to an unpublished Stanford technical report, "A procedure for determining the loaded face of a die," 1952, written by S. G. Allen, Jr., which reports some results of H. Rubin and the late M. A. Girshick. They were concerned with obtaining a large-sample approximation for (3) when (1) holds. The approximation that they propose is different from ours and appears to be much more tedious to apply. We have not attempted to determine which approximation is the better.

For arbitrary k we let

$$W_i = \frac{Y_{(k)N}}{N} - \frac{Y_{(i)N}}{N} \qquad (i = 1, 2, \cdots, k-1) \tag{8}$$

and approximate (3) by

$$\tilde{\phi}_k = \Pr\{W_1 \geqq 0, W_2 \geqq 0, \cdots, W_{k-1} \geqq 0\}. \tag{9}$$

TABLE A-2

Exact Probability of a Correct Selection for $k = 2$ and selected θ^* and N when $p_{[2]}/p_{[1]} = \theta^*$

N	θ^*									
	1.02	1.04	1.06	1.08	1.10	1.20	1.30	1.40	1.50	1.60
1	.504951	.509804	.514563	.519231	.523810	.545455	.565217	.583333	.600000	.615385
2	.504951	.509804	.514563	.519231	.523810	.545455	.565217	.583333	.600000	.615385
3	.507426	.514704	.521838	.528832	.535687	.567994	.597271	.623843	.648000	.670005
4	.507426	.514704	.521838	.528832	.535687	.567994	.597271	.623843	.648000	.670005
5	.509282	.518378	.527290	.536022	.544575	.584759	.620903	.653381	.682560	.708788
6	.509282	.518378	.527290	.536022	.544575	.584759	.620903	.653381	.682560	.708788
7	.510828	.521438	.531830	.542005	.551965	.598614	.640261	.677312	.710208	.739386
8	.510828	.521438	.531830	.542005	.551965	.598614	.640261	.677312	.710208	.739386
9	.512181	.524114	.535798	.547232	.558417	.610637	.656910	.697670	.733432	.764734
10	.512181	.524114	.535798	.547232	.558417	.610637	.656910	.697670	.733432	.764734
11	.513399	.526522	.539367	.551930	.564210	.621369	.671640	.715483	.753498	.786332
12	.513399	.526522	.539367	.551930	.564210	.621369	.671640	.715483	.753498	.786332
13	.514515	.528729	.542636	.556230	.569509	.631125	.684913	.731359	.771156	.805076
14	.514515	.528729	.542636	.556230	.569509	.631125	.684913	.731359	.771156	.805076
15	.515551	.530777	.545668	.560217	.574417	.640109	.697028	.745691	.786897	.821554
16	.515551	.530777	.545668	.560217	.574417	.640109	.697028	.745691	.786897	.821554
17	.516523	.532697	.548509	.563949	.579009	.648462	.708193	.758754	.801064	.836180
18	.516523	.532697	.548509	.563949	.579009	.648462	.708193	.758754	.801064	.836180
19	.517440	.534509	.551189	.567468	.583336	.656286	.718558	.770748	.813908	.849257
20	.517440	.534509	.551189	.567468	.583336	.656286	.718558	.770748	.813908	.849257
21	.518312	.536229	.553734	.570807	.587437	.663657	.728237	.781826	.825622	.861019
22	.518312	.536229	.553734	.570807	.587437	.663657	.728237	.781826	.825622	.861019
23	.519143	.537871	.556160	.573989	.591342	.670635	.737319	.792107	.836357	.871648
24	.519143	.537871	.556160	.573989	.591342	.670635	.737319	.792107	.836357	.871648
25	.519940	.539444	.558484	.577034	.595077	.677267	.745874	.801686	.846232	.881292
26	.519940	.539444	.558484	.577034	.595077	.677267	.745874	.801686	.846232	.881292
27	.520706	.540956	.560716	.579958	.598660	.683590	.753961	.810641	.855348	.890072
28	.520706	.540956	.560716	.579958	.598660	.683590	.753961	.810641	.855348	.890072
29	.521445	.542413	.562867	.582773	.602107	.689639	.761626	.819036	.863787	.898087
30	.521445	.542413	.562867	.582773	.602107	.689639	.761626	.819036	.863787	.898087

TABLE A-2 (*Continued*)

	1.70	1.80	1.90	2.00	2.20	2.40	2.60	2.80	3.00	10.00
1	.629630	.642857	.655172	.666667	.687500	.705882	.722222	.736842	.750000	.909091
2	.629630	.642857	.655172	.666667	.687500	.705882	.722222	.736842	.750000	.909091
3	.690088	.708455	.725286	.740741	.768066	.791370	.811385	.828692	.843750	.976709
4	.690088	.708455	.725286	.740741	.768066	.791370	.811385	.828692	.843750	.976709
5	.732384	.753637	.772806	.790123	.819994	.844615	.865048	.882123	.896484	.993474
6	.732384	.753637	.772806	.790123	.819994	.844615	.865048	.882123	.896484	.993474
7	.765261	.788215	.808593	.826703	.857182	.881462	.900934	.916658	.929443	.998093
8	.765261	.788215	.808593	.826703	.857182	.881462	.900934	.916658	.929443	.998093
9	.792095	.816001	.836890	.855154	.885145	.908237	.926132	.940096	.951073	.999428
10	.792095	.816001	.836890	.855154	.885145	.908237	.926132	.940096	.951073	.999428
11	.814623	.838967	.859905	.877915	.906773	.928249	.944330	.956457	.965673	.999826
12	.814623	.838967	.859905	.877915	.906773	.928249	.944330	.956457	.965673	.999826
13	.833885	.858301	.878969	.896461	.923811	.943483	.957717	.968090	.975710	.999946
14	.833885	.858301	.878969	.896461	.923811	.943483	.957717	.968090	.975710	.999946
15	.850569	.874788	.894967	.911768	.937407	.955231	.967692	.976468	.982700	.999983
16	.850569	.874788	.894967	.911768	.937407	.955231	.967692	.976468	.982700	.999983
17	.865159	.888983	.908521	.924525	.948360	.964376	.975196	.982560	.987615	.999995
18	.865159	.888983	.908521	.924525	.948360	.964376	.975196	.982560	.987615	.999995
19	.878012	.901294	.920089	.935234	.957251	.971550	.980883	.987023	.991097	.999998
20	.878012	.901294	.920089	.935234	.957251	.971550	.980883	.987023	.991097	.999998
21	.889402	.912036	.930019	.944277	.964509	.977209	.985219	.990311	.993577	.999999
22	.889402	.912036	.930019	.944277	.964509	.977209	.985219	.990311	.993577	.999999
23	.899543	.921452	.938586	.951950	.970463	.981695	.988540	.992745	.995353	1.000000
24	.899543	.921452	.938586	.951950	.970463	.981695	.988540	.992745	.995353	1.000000
25	.908609	.929740	.946005	.958486	.975366	.985265	.991094	.994555	.996630	1.000000
26	.908609	.929740	.946005	.958486	.975366	.985265	.991094	.994555	.996630	1.000000
27	.916740	.937058	.952450	.964073	.979418	.988116	.993066	.995904	.997549	1.000000
28	.916740	.937058	.952450	.964073	.979418	.988116	.993066	.995904	.997549	1.000000
29	.924053	.943538	.958069	.968861	.982776	.990399	.994590	.996914	.998216	1.000000
30	.924053	.943538	.958069	.968861	.982776	.990399	.994590	.996914	.998216	1.000000

6.1 The normal approximation. In the least favorable configuration the chance variables $W_i\,(i = 1, 2, \cdots, k-1)$ have a $(k-1)$-variate distribution with

$$E\{W_i\} = \frac{\theta^* - 1}{\theta^* + k - 1} \tag{10a}$$

$$\text{Var}\,\{W_i\} = \frac{(k+2)\theta^* + (k-2)}{N(\theta^* + k - 1)^2} \tag{10b}$$

$$\underset{i \neq j}{\text{Cov}}\,\{W_i, W_j\} = \frac{(k+1)\theta^* - 1}{N(\theta^* + k - 1)^2} \tag{10c}$$

$$\text{(10d)} \qquad \operatorname{Corr}_{i \neq j} \{W_i, W_j\} = \frac{(k+1)\theta^* - 1}{(k+2)\theta^* + (k-2)};$$

it is to be noted that the W_i have a common mean and a common variance, and all pairs (W_i, W_j) with $i \neq j$ have a common correlation. The standardized variables obtained by subtracting the common mean from each W_i and dividing the differences by the common standard deviation have zero mean, unit variance and correlation (10d). As N approaches infinity it can be shown that their joint

TABLE A-3

Exact Probability of a Correct Selection for $k = 3$ *and selected* θ^* *and* N *when* $p_{[3]}/p_{[2]} = p_{[3]}/p_{[1]} = \theta^*$

N	θ^*									
	1.02	1.04	1.06	1.08	1.10	1.20	1.30	1.40	1.50	1.60
1	.337748	.342105	.346405	.350649	.354839	.375000	.393939	.411765	.428571	.444444
2	.337748	.342105	.346405	.350649	.354839	.375000	.393939	.411765	.428571	.444444
3	.339230	.345067	.350845	.356563	.362223	.389648	.415644	.440261	.463557	.485597
4	.340211	.347015	.353746	.360404	.366988	.398804	.428798	.457023	.483549	.508459
5	.340708	.348015	.355254	.362424	.369524	.403954	.436571	.467376	.496400	.523701
6	.341530	.349659	.357717	.365703	.373614	.412000	.448349	.482601	.514760	.544870
7	.342261	.351116	.359896	.368599	.377220	.419041	.458579	.495737	.530497	.562902
8	.342690	.351980	.361198	.370341	.379405	.423441	.465140	.504341	.540988	.575098
9	.343331	.353264	.363125	.372909	.382612	.429767	.474385	.516236	.555226	.591360
10	.343924	.354448	.364899	.375270	.385556	.435526	.482730	.526881	.567859	.605664
11	.344311	.355228	.366075	.376845	.387531	.439500	.488625	.534549	.577108	.616279
12	.344852	.356312	.367703	.379016	.390244	.444851	.496410	.544485	.588874	.629544
13	.345363	.357335	.369237	.381060	.392794	.449841	.503615	.553613	.599602	.641549
14	.345718	.358049	.370316	.382506	.394608	.453486	.508992	.560545	.607866	.650901
15	.346193	.359004	.371751	.384422	.397003	.458204	.515819	.569185	.617984	.662159
16	.346649	.359916	.373121	.386247	.399282	.462662	.522228	.577239	.627350	.672506
17	.346978	.360581	.374125	.387593	.400971	.466050	.527195	.583582	.634822	.680846
18	.347408	.361444	.375422	.389326	.403137	.470311	.533327	.591273	.643725	.690617
19	.347822	.362275	.376671	.390991	.405217	.474375	.539137	.598513	.652051	.699693
20	.348131	.362899	.377614	.392256	.406805	.477552	.543766	.604368	.658865	.707190
21	.348525	.363692	.378808	.393850	.408798	.481465	.549362	.611321	.666818	.715798
22	.348908	.364459	.379962	.395391	.410723	.485220	.554699	.617912	.674310	.723854
23	.349200	.365050	.380855	.396589	.412226	.488221	.559043	.623353	.680564	.730636
24	.349566	.365787	.381966	.398073	.414083	.491856	.564207	.629708	.687746	.738300
25	.349923	.366505	.383045	.399514	.415882	.495360	.569158	.635764	.694548	.745512
26	.350202	.367067	.383896	.400655	.417315	.498211	.573255	.640847	.700315	.751678
27	.350546	.367759	.384939	.402049	.419059	.501616	.578062	.646703	.706852	.758554
28	.350882	.368435	.385956	.403407	.420755	.504912	.582687	.652307	.713070	.765054
29	.351147	.368973	.386770	.404499	.422126	.507632	.586569	.657072	.718410	.770679
30	.351472	.369627	.387756	.405818	.423776	.510844	.591073	.662503	.724397	.776883

TABLE A-3 (*Continued*)

	1.70	1.80	1.90	2.00	2.20	2.40	2.60	2.80	3.00	10.00
1	.459459	.473684	.487179	.500000	.523810	.545455	.565217	.583333	.600000	.833333
2	.459459	.473684	.487179	.500000	.523810	.545455	.565217	.583333	.600000	.833333
3	.506446	.526170	.544834	.562500	.595076	.624343	.650694	.647479	.696000	.937500
4	.531844	.553794	.574400	.593750	.629013	.660201	.687858	.712457	.734400	.954861
5	.549348	.573422	.596006	.617188	.655677	.689539	.719367	.745687	.768960	.976562
6	.573002	.599248	.623707	.646484	.687420	.722879	.753615	.780302	.803520	.985605
7	.593033	.620996	.646908	.670898	.713621	.750156	.781411	.808186	.831168	.992135
8	.606741	.636019	.663056	.687988	.732096	.769443	.801050	.827817	.850522	.995066
9	.624709	.655391	.383552	.709351	.754530	.792237	.823677	.849902	.871811	.997263
10	.640379	.672142	.701122	.727509	.773277	.810967	.841960	.867454	.888455	.998342
11	.652138	.684827	.714528	.741447	.787793	.825541	.856218	.881142	.901414	.999029
12	.666581	.700148	.730459	.757750	.804259	.841598	.871496	.895425	.914596	.999425
13	.679553	.713808	.744556	.772069	.818512	.855299	.884350	.907277	.925387	.999662
14	.689754	.724627	.755785	.783525	.829974	.866334	.894690	.916778	.933995	.999797
15	.701829	.737229	.768662	.796463	.842547	.878110	.905440	.926415	.942521	.999881
16	.712849	.748649	.780247	.808021	.853620	.888334	.914642	.934550	.949621	.999929
17	.721803	.757985	.789763	.817546	.862776	.896789	.922234	.941231	.955417	.999957
18	.732124	.768579	.800397	.828033	.872575	.905597	.929943	.947853	.961031	.999975
19	.741646	.778285	.810073	.837510	.881309	.913341	.936627	.953517	.965769	.999985
20	.749569	.786405	.818200	.845490	.888677	.919865	.942239	.958246	.969696	.999991
21	.758521	.795439	.827106	.854109	.896418	.926542	.947841	.962856	.973441	.999995
22	.766844	.803783	.835277	.861964	.903378	.932464	.952742	.966836	.976630	.999997
23	.773897	.810887	.842256	.868687	.909339	.937525	.956912	.970200	.979306	.999998
24	.781741	.818669	.849790	.875841	.915513	.942633	.961019	.973439	.981825	.999999
25	.789077	.825899	.856745	.882401	.921100	.947194	.964638	.976255	.983987	.999999
26	.795385	.832141	.862765	.888089	.925944	.951138	.967749	.978658	.985817	1.000000
27	.802311	.838895	.869184	.894068	.930902	.955073	.970782	.980949	.987524	1.000000
28	.808819	.845199	.875139	.899580	.935412	.958605	.973468	.982950	.988996	1.000000
29	.814481	.850706	.880351	.904409	.939360	.961685	.975795	.984672	.990250	1.000000
30	.820632	.856601	.885853	.909437	.943363	.964731	.978045	.986299	.991412	1.000000

distribution approaches a $(k-1)$-variate normal distribution. Thus, for large N the probability of a correct selection in the least favorable configuration can be approximated by the volume under a $(k-1)$-variate normal surface. However for $k \geqq 4$ such volumes are tabulated (see Table 1 of [1], Table A1 of [5], and [9]) only for the particular correlation matrix $\{\rho_{ij}\}$ where

$$\rho_{ij} = \begin{cases} 1 & \text{if } i = j \\ \frac{1}{2} & \text{if } i \neq j. \end{cases} \tag{11}$$

Although the matrix of Corr $\{W_i, W_j\}$ *is* of this *form*, the correlation coefficient when $i \neq j$ is *not* $\frac{1}{2}$; in fact, as k approaches infinity this correlation coefficient approaches $\theta^*/(\theta^* + 1)$. Because of the lack of appropriate tables this approach

was abandoned. The approach described in the next section yielded the desired results.

6.2. The arcsin transformation and the normal approximation.

6.2.1. Derivation of formulae for the approximation.

We next consider the chance variables

TABLE A-4

Exact Probability of a Correct Selection for $k = 4$ and selected θ^* and N when $p_{[4]}/p_{[3]} = p_{[4]}/p_{[2]} = p_{[4]}/p_{[1]} = \theta^*$

N	θ^*									
	1.02	1.04	1.06	1.08	1.10	1.20	1.30	1.40	1.50	1.60
1	.253731	.257426	.261084	.264706	.268293	.285714	.302326	.318182	.333333	.347826
2	.253731	.257426	.261084	.264706	.268293	.285714	.302326	.318182	.333333	.347826
3	.254673	.259318	.263935	.268522	.273081	.295432	.317041	.337904	.358025	.377414
4	.255611	.261192	.266744	.272264	.277752	.304688	.330730	.355833	.379973	.403143
5	.256139	.262250	.268331	.274382	.280401	.309977	.338609	.366224	.392775	.418245
6	.256583	.263145	.269686	.276203	.282694	.314699	.345828	.375948	.404969	.432836
7	.257172	.264331	.271472	.278595	.285694	.320770	.354949	.388029	.419872	.450388
8	.257687	.265365	.273030	.280677	.288304	.326017	.362783	.398342	.432516	.465190
9	.258099	.266195	.274282	.282357	.290415	.330318	.369274	.406967	.443171	.477740
10	.258492	.266989	.275485	.283975	.292454	.334516	.375660	.415493	.453734	.490198
11	.258928	.267868	.276814	.285758	.294695	.339075	.382512	.424536	.464814	.503121
12	.259327	.268673	.278029	.287388	.296744	.343240	.388764	.432773	.474882	.514835
13	.259681	.269389	.279113	.288846	.298580	.347007	.394460	.440317	.484138	.525629
14	.260022	.270078	.280158	.290253	.300355	.350670	.400013	.447681	.493169	.536145
15	.260383	.270808	.281264	.291741	.302228	.354499	.405770	.455249	.502373	.546771
16	.260724	.271497	.282308	.293144	.303995	.358115	.411204	.462384	.511035	.556749
17	.261038	.272133	.283272	.294443	305634	.361490	.416297	.469092	.519189	.566146
18	.261340	.272747	.284205	.295701	.307222	.364774	.421263	.475631	.527129	.575274
19	.261657	.273387	.285177	.297011	.308874	.368164	.426353	.482287	.535152	.584432
20	.261959	.274001	.286109	.298266	.310457	.371415	.431231	.488657	.542817	.593158
21	.262243	.274578	.286985	.299447	.311949	.374495	.435869	.494723	.550117	.601462
22	.262519	.275137	.287836	.300597	.313402	.377504	.440402	.500650	.557238	.609543
23	.262804	.275716	.288716	.301784	.314900	.380587	.445020	.506648	.564399	.617614
24	.263079	.276276	.289567	.302931	.316349	.383569	.449483	.512437	.571294	.625364
25	.263341	.276807	.290375	.304023	.317730	.386423	.453764	.517995	.577911	.632792
26	.263595	.277324	.291164	.305090	.319079	.389219	.457960	.523439	.584380	.640033
27	.263857	.277857	.291975	.306185	.320463	.392071	.462216	.528926	.590860	.647242
28	.264112	.278375	.292763	.307249	.321808	.394844	.466352	.534250	.597133	.654198
29	.264355	.278870	.293518	.308270	.323100	.397516	.470344	.539390	.603185	.660898
30	.264592	.279355	.294257	.309271	.324366	.400140	.474265	.544434	.609112	.667440

TABLE A-4 (*Continued*)

	1.70	1.80	1.90	2.00	2.20	2.40	2.60	2.80	3.00	10.00
1	.361702	.375000	.387755	.400000	.423077	.444444	.464286	.482759	.500000	.769231
2	.361702	.375000	.387755	.400000	.423077	.444444	.464286	.482759	.500000	.769231
3	.396088	.414063	.431359	.448000	.479404	.508459	.535350	.560253	.583333	.892126
4	.425352	.446615	.466955	.486400	.522732	.555877	.586110	.613697	.638889	.929939
5	.442630	.465942	.488203	.509440	.548979	.584855	.617382	.646869	.673611	.956118
6	.459521	.485016	.509329	.532480	.575418	.614126	.648945	.680232	.708333	.972900
7	.479526	.507268	.533616	.558592	.604569	.645564	.682011	.714362	.743056	.983657
8	.496297	.525812	.553736	.580096	.628297	.670860	.708314	.741205	.770062	.989681
9	.510588	.541676	.571001	.598589	.648751	.692668	.730951	.764235	.793130	.993584
10	.524771	.557400	.588076	.616825	.668762	.713796	.752642	.786038	.814697	.996024
11	.539326	.573366	.605234	.634964	.688278	.734014	.773022	.806167	.834270	.997522
12	.552481	.587752	.620642	.651193	.705602	.751811	.790802	.823564	.851026	.998450
13	.564618	.601028	.634853	.666144	.721496	.768040	.806895	.839178	.865925	.999033
14	.576412	.613886	.648563	.680503	.736604	.783293	.821837	.853493	.879409	.999395
15	.588232	.626666	.662079	.694543	.751145	.797744	.835780	.866653	.891625	.999622
16	.599300	.638596	.674651	.707554	.764506	.810902	.848351	.878395	.902410	.999763
17	.609715	.649804	.686436	.719716	.776908	.823012	.859809	.888988	.912033	.999851
18	.619802	.660620	.697760	.731348	.788646	.834343	.870404	.898662	.920712	.999907
19	.629849	.671313	.708875	.742683	.799919	.845071	.880292	.907565	.928587	.999942
20	.639392	.681436	.719356	.753325	.810407	.854951	.889302	.915588	.935605	.999963
21	.648459	.691030	.729261	.763347	.820203	.864092	.897552	.922853	.941887	.999977
22	.657254	.700300	.738788	.772941	.829482	.872652	.905187	.929494	.947559	.999986
23	.665978	.709434	.748111	.782265	.838375	.880744	.912303	.935600	.952702	.999991
24	.674328	.718142	.756962	.791078	.846700	.888239	.918823	.941130	.957307	.999994
25	.682313	.726446	.765375	.799421	.854511	.895201	.924814	.946155	.961441	.999996
26	.690070	.734481	.773477	.807418	.861918	.901727	.930363	.950752	.965176	.999997
27	.697743	.742376	.781387	.815172	.869005	.907887	.935530	.954974	.968560	.999998
28	.705121	.749939	.788930	.822533	.875664	.913614	.940279	.958810	.971600	.999999
29	.712210	.757181	.796128	.829528	.881935	.918950	.944656	.962304	.974335	1.000000
30	.719106	.764196	.803068	.836237	.887883	.923956	.948710	.965501	.976808	1.000000

$$(12) \qquad Z_i = 2 \arcsin \sqrt{\frac{Y_{(i)N}}{N}} - 2 \arcsin \sqrt{\frac{Y_{(i)N}}{N}} \quad (i = 1, 2, \cdots, k-1),$$

and for arbitrary k we shall write (9) as

$$(13) \qquad \phi_k = \Pr\{Z_1 \geqq 0, Z_2 \geqq 0, \cdots, Z_{k-1} \geqq 0\}.$$

The chance variables Z_i $(i = 1, 2, \cdots, k-1)$ also have a $(k-1)$-variate distribution, with means, variances, and covariances which can be expressed as power series in $1/N$. In the l.f.c. these power series involve only k, N and θ^*. They are found in the following way. We expand 2 arcsin $(Y_{(i)N}/N)^{1/2}$, consid-

ered as a function of $Y_{(i)N}/N$, in a Taylor series around the point $Y_{(i)N}/N = p_{[i]}$ obtaining

$$
\begin{aligned}
2 \arcsin \sqrt{\frac{Y_{(i)N}}{N}} &= f(p_{[i]}) + f'(p_{[i]}) \left(\frac{Y_{(i)N}}{N} - p_{[i]} \right) \\
&\quad + \frac{1}{2} f''(p_{[i]}) \left(\frac{Y_{(i)N}}{N} - p_{[i]} \right)^2 + \frac{1}{6} f'''(p_{[i]}) \left(\frac{Y_{(i)N}}{N} - p_{[i]} \right)^3 \\
&\quad + \frac{1}{24} f^{iv}(p_{[i]}) \left(\frac{Y_{(i)N}}{N} - p_{[i]} \right)^4 + O\left(\frac{1}{N^5} \right)
\end{aligned} \tag{14}
$$

where

$$
\begin{aligned}
f(p_{[i]}) &= 2 \arcsin \sqrt{p_{[i]}} \\
f'(p_{[i]}) &= \frac{1}{\sqrt{p_{[i]}(1 - p_{[i]})}} \\
f''(p_{[i]}) &= \frac{2p_{[i]} - 1}{2\sqrt{p_{[i]}^3 (1 - p_{[i]})^3}} \\
f'''(p_{[i]}) &= \frac{8p_{[i]}^2 - 8p_{[i]} + 3}{4\sqrt{p_{[i]}^5 (1 - p_{[i]})^5}} \\
f^{iv}(p_{[i]}) &= \frac{3(2p_{[i]} - 1)(8p_{[i]}^2 - 8p_{[i]} + 5)}{8\sqrt{p_{[i]}^7 (1 - p_{[i]})^7}}.
\end{aligned} \tag{15}
$$

Now $E\{Z_i\}$, Var $\{Z_i\}$, and Cov $\{Z_i, Z_j\}$ all can be expressed in terms of $E\{2 \arcsin (Y_{(i)N}/N)^{1/2}\}$, $E\{[2 \arcsin (Y_{(i)N}/N)^{1/2}]^2\}$ and $E\{[2 \arcsin (Y_{(i)N}/N)^{1/2}] [2 \arcsin (Y_{(j)N}/N)^{1/2}]\}$, and these latter involve the moments of multinomially distributed variables which are well known. Thus, for example,

$$
\begin{aligned}
E\{2 \arcsin \sqrt{Y_{(i)N}/N}\} &= f(p_{[i]}) + f''(p_{[i]}) \frac{p_{[i]}(1 - p_{[i]})}{2N} \\
&\quad + f'''(p_{[i]}) \frac{p_{[i]}(1 - p_{[i]})(1 - 2p_{[i]})}{6N^2} \\
&\quad + f^{iv}(p_{[i]}) \left[\frac{p_{[i]}^2 (1 - p_{[i]})^2}{8N^2} + \frac{p_{[i]}(1 - p_{[i]})(1 - 6p_{[i]} + 6p_{[i]}^2)}{24N^3} \right] \\
&\quad + O(1/N^4) \qquad (i = 1, 2, \cdots, k - 1)
\end{aligned} \tag{16}
$$

which immediately yields $E\{Z_i\}$ up to terms of order $1/N^4$.

In the l.f.c. we use (2) and obtain after simplifications

$$
E\{Z_i\} = a_1 + \frac{a_2}{N} + O(1/N^2) \qquad (i = 1, 2, \cdots, k - 1) \tag{17a}
$$

$$
\mathrm{Var}\{Z_i\} = \frac{b_1}{N} + \frac{b_2}{N^2} + O(1/N^3) \tag{17b}
$$

where

$$(18a)\quad a_1 = 2\arcsin\sqrt{\frac{\theta^*}{\theta^*+k-1}} - 2\arcsin\sqrt{\frac{1}{\theta^*+k-1}}$$

$$(18b)\quad a_2 = \frac{\theta^*+k-3}{4\sqrt{\theta^*+k-2}} + \frac{\theta^*-k+1}{4\sqrt{(k-1)\theta^*}}$$

$$(18c)\quad b_1 = 2 + 2\sqrt{\frac{\theta^*}{(k-1)(\theta^*+k-2)}}$$

$$(18d)\quad b_2 = \frac{(\theta^*+k-1)^2(k\theta^*+k-2) - \theta^*(k-\theta^*-1)(\theta^*+k-3)}{4\sqrt{\theta^*(k-1)^3(\theta^*+k-2)^3}} + \frac{3(\theta^*+k-1)^2(k\theta^*+k-2)}{8\theta^*(k-1)(\theta^*+k-2)} - 1.$$

In addition,

$$(19)\quad \operatorname{Corr}_{i\neq j}\{Z_i, Z_j\} = \frac{1+2\sqrt{\frac{\theta^*}{(k-1)(\theta^*+k-2)}} - \frac{1}{\theta^*+k-2} + \frac{c}{N} + O\left(\frac{1}{N^2}\right)}{2+2\sqrt{\frac{\theta^*}{(k-1)(\theta^*+k-2)}} + \frac{b_2}{N} + O\left(\frac{1}{N^2}\right)}$$

where

$$(20)\quad c = \frac{(\theta^*+k-1)^2(k\theta^*+k-2) + \theta^*(\theta^*-k+1)(\theta^*+k-3)}{4\sqrt{\theta^*(k-1)^3(\theta^*+k-2)^3}} - \frac{2(\theta^*+k-1)^2(\theta^*+k-2) - (\theta^*+k-3)^2}{8(\theta^*+k-2)^3}.$$

The following table shows how (19) varies with k and θ^* when N is large. (In these computations, N in (19) is assumed to be infinite.)

θ^*	k					
	3	4	5	6	10	36
1.10	0.512	0.508	0.506	0.505	0.503	0.501
1.25	0.527	0.519	0.514	0.511	0.506	0.502
1.50	0.548	0.533	0.526	0.521	0.512	0.503
1.75	0.564	0.546	0.536	0.529	0.517	0.505
2.00	0.577	0.556	0.544	0.536	0.521	0.506
10.00	0.674	0.645	0.626	0.611	0.577	0.527

It is seen from (19) with $N = +\infty$ (and from the above table) that as θ^* approaches unity, (19) approaches $\frac{1}{2}$ for any k; also, as k approaches infinity, (19) approaches $\frac{1}{2}$ for any θ^*.

As with the W_i, when (1) holds the Z_i have a common mean and a common variance, and all pairs (Z_i, Z_j) with $i \neq j$ have a common correlation. Also as N approaches infinity it can be shown that the joint distribution of the standardized variables approaches a $(k-1)$-variate normal distribution with zero means, unit variances, and a correlation matrix of the form of (11); *and the common correlation coefficient* (19) *for* $i \neq j$ *is approximately* $\frac{1}{2}$. Thus, the tables in [1], [5], and [9] can be used for finding the volume under the $(k-1)$-variate normal surface, and hence an approximation to the probability of a correct selection can be obtained.

6.2.2. The approximation. To obtain the approximation we let

$$A = a_1 + \frac{a_2}{N} \tag{21a}$$

and

$$B = \frac{b_1}{N} + \frac{b_2}{N^2}. \tag{21b}$$

Then

$$\phi_k \sim \int_{\frac{A}{\sqrt{B}}}^{\infty} \int_{\frac{A}{\sqrt{B}}}^{\infty} \cdots \int_{\frac{A}{\sqrt{B}}}^{\infty} g(t_1, t_2, \cdots, t_{k-1})\, dt_1\, dt_2 \cdots dt_{k-1}, \tag{22}$$

where $g(t_1, t_2, \cdots, t_{k-1})$ is the $(k-1)$-variate normal density function with zero means, unit variances, and correlation matrix (11). Since [9] gives $P_k(\Lambda)$ as a function of Λ where

$$P_k(\Lambda) = \int_{\frac{\Lambda}{\sqrt{2}}}^{\infty} \int_{\frac{\Lambda}{\sqrt{2}}}^{\infty} \cdots \int_{\frac{\Lambda}{\sqrt{2}}}^{\infty} g(t_1, t_2, \cdots, t_{k-1})\, dt_1\, dt_2, \cdots, dt_{k-1}, \tag{23}$$

we obtain

$$\phi_k \sim P_k\left(A\sqrt{\frac{2}{B}}\right). \tag{24}$$

6.2.3. Tables based on the approximation. Some computations were made to indicate the goodness of the approximation. These approximate probabilities are listed along with the corresponding exact probabilities (which were extracted from Tables A-2, A-3, and A-4) in Tables B-2, B-3, and B-4. They were computed as follows: The quantities a_1, a_2, b_1, and b_2 were computed for each k and θ^* using (18a), (18b), (18c), and (18d), respectively; then A and B were computed for each N using (21a) and (21b), respectively; then $\Lambda = A(2/B)^{1/2}$

TABLE B-2

Exact[1] and Approximate[1] Probability of a Correct Selection for $k = 2$ when $p_{[2]}/p_{[1]} = \theta^*$

θ^*	$N = 10$	$N = 15$	$N = 20$	$N = 25$	$N = 30$
1.02	.51250	.51530	.51767	.51974	.52163
	.512181	.515551	.517440	.519940	.521445
1.10	.55993	.57322	.58437	.59414	.60293
	.558417	.574417	.583336	.595077	.602107
1.50	.73764	.78204	.81587	.84279	.86481
	.733432	.786897	.813908	.846232	.863787
2.00	.85744	.90506	.93711	.95494	.96836
	.855154	.911768	.935234	.958486	.968861
3.00	.94834	.97750	.98985	.99533	.99781
	.951073	.982700	.991097	.996630	.998216
10.00	.99700	.99973	.99998	1.00000	1.00000
	.999428	.999983	.999998	1.000000	1.000000

[1] Exact probabilities are given to six decimal places; approximate probabilities are given to five decimal places.

TABLE B-3

Exact[1] and Approximate[1] Probability of a Correct Selection for $k = 3$ when $p_{[1]}/p_{[2]} = p_{[2]}/p_{[3]} = \theta^*$

θ^*	$N = 10$	$N = 15$	$N = 20$	$N = 25$	$N = 30$
1.02	.34361	.34593	.34789	.34963	.35120
	.343924	.346193	.348131	.349923	.351472
1.10	.38408	.39577	.40571	.41451	.42251
	.385556	.397003	.406805	.415882	.423776
1.50	.56255	.61374	.65541	.69061	.72097
	.567859	.617984	.658865	.694548	.724398
2.00	.71977	.79077	.84134	.87854	.90642
	.727509	.796463	.845490	.882401	.909437
3.00	.87858	.93681	.96644	.98196	.99022
	.888455	.942521	.969696	.983987	.991412
10.00	.99450	.99958	.99995	1.00000	1.00000
	.998342	.999881	.999991	.999999	1.000000

[1] Exact probabilities are given to six decimal places; approximate probabilities are given to five decimal places.

was computed; then [9] was entered with Λ and k, and $P_A(\Lambda)$ was read out. (See section 6.2.4 for a description of the tables in [1], [5], and [9].) Comparison of the approximate and exact probabilities in Tables B-2, B-3, and B-4 indicates that the approximation is excellent even when N is only moderately large. Of course, the approximation breaks down if θ^* is large and k is small for then (19) differs too much from $\frac{1}{k}$.

TABLE B-4

Exact[1] *and Approximate*[1] *Probability of a Correct Selection for* $k = 4$ *when* $p_{[4]}/p_{[3]} = p_{[4]}/p_{[2]} = p_{[4]}/p_{[1]} = \theta^*$

θ^*	$N = 10$	$N = 15$	$N = 20$	$N = 25$	$N = 30$
1.02	.25812	.25997	.26154	.26292	.26315
	.258492	.260383	.261959	.263341	.264592
1.10	.29066	.30022	.30839	.31567	.32232
	.292454	.302228	.310457	.317730	.324366
1.50	.44645	.49462	.53529	.57078	.60235
	.453734	.502373	.542817	.577911	.609112
2.00	.60566	.68421	.74450	.79203	.83003
	.616825	.694543	.753325	.799421	.836237
3.00	.80045	.88202	.92943	.95751	.97430
	.814697	.891625	.935605	.961441	.976808
10.00	.98975	.99878	.99986	1.00000	1.00000
	.996024	.999622	.999963	.999997	1.000000

[1] Exact probabilities are given to six decimal places; approximate probabilities are given to five decimal places.

The problem of including a correction for continuity in the approximation for small N was considered, but the authors were not successful in finding one which would give uniform improvements over the approximation finally adopted.

6.2.4. Reference tables. Since [9], Table A1 of [5], and Table 1 of [1] employ different notations, some comments about these tables might be appropriate.

In [9], $P_k(\Lambda)$ is tabulated as a function of Λ for various k. In the notation of these tables, $\Lambda = x$ and $P_k(\Lambda) = P(1, k)$ which is a function of Λ. The tabulations are for $\Lambda = 0.00\ (0.01)\ 6.09$ and $k = 2(1)10$. These tables were originally computed for the purpose of preparing Table 1 of [1].

In Table 1 of [1], Λ is tabulated as a function of k and $P_k(\Lambda)$. In the notation of these tables, $\Lambda = N^{1/2}\lambda$, and the columns to be entered are those headed $t = 1$. The tabulations are for $P_k(\Lambda) = 0.10\ (0.05)\ 0.80\ (0.02)\ 0.90\ (0.01)\ 0.99$, 0.9950, 0.9990, 0.9995 and $k = 2(1)10$.

In Table A1 of [5], Λ is tabulated as a function of k and $P_k(\Lambda)$. In the notation of these tables $\Lambda = u_\alpha(n)$, $k = n + 1$, and $P_k(\Lambda) = 1 - \alpha$. The tabulations are for $P_k(\Lambda) = 0.75$, 0.90, 0.95, 0.975, 0.99 and $k = 2(1)51$.

7. Choice of N to meet the requirement. We now shall show how to determine N^*, the smallest N which will guarantee the requirement. To guarantee the requirement in the least favorable configuration we must have

$$\int_{\frac{-\Lambda}{\sqrt{B}}}^{\infty} \int_{\frac{-\Lambda}{\sqrt{B}}}^{\infty} \cdots \int_{\frac{-\Lambda}{\sqrt{B}}}^{\infty} g(t_1, t_2, \cdots, t_{k-1})\, dt_1\, dt_2 \cdots dt_{k-1} \geqq P^*. \tag{25}$$

Hence, N must be chosen large enough to make

$$\frac{A}{\sqrt{B}} = \frac{a_1 + a_2/N}{\sqrt{b_1/N + b_2/N^2}} \geqq \frac{\Lambda}{\sqrt{2}} \tag{26}$$

where Λ is determined from [1], [5], or [9] to satisfy the equation

$$P_k(\Lambda) = P^*. \tag{27}$$

We consider the inequality in (26) and note that the middle expression approaches infinity as N grows large. Therefore, since a_1 is positive there exists (for any a_1, a_2, b_1, b_2, and Λ) a smallest integer, which we denote by N^*, with the property that the inequality (26) holds. When N is large, the terms in (26) involving a_2 and b_2 can be ignored, and N^* is approximately the smallest integer equal to or greater than

$$\frac{\Lambda^2 b_1}{2a_1^2}. \tag{28}$$

A tendency for the approximation (22) to underestimate the exact value of (3) (except for $k = 2$ when N is *even*) is evident from examination of Tables B-1, B-2, and B-3. Hence, one usually can expect the value of N^* obtained by the methods described above to err on the conservative side, i.e., to be somewhat larger than the exact value of N^* which is required.

8. Numerical example. Suppose that one were interested in selecting that one of the thirty-six bettable numbers on an unbalanced roulette wheel which has the largest probability associated with it. Suppose further that he specifies that if this probability is at least 10% larger than the second-largest probability, he wishes to make a correct selection with probability at least 0.90. Then we have $k = 36$, $\theta^* = 1.10$, and $P^* = 0.90$; the least favorable configuration is

$$\begin{aligned} p_{[1]} = \cdots = p_{[35]} &= \frac{1}{36.1} = 0.02770083 \\ p_{[36]} &= \frac{1.1}{36.1} = 0.03047091. \end{aligned} \tag{29}$$

We can anticipate here that N^* will be large, and hence we need compute only

$$a_1 = 2 \arcsin\sqrt{0.03047091} - 2 \arcsin\sqrt{0.02770083} = 0.0164885 \tag{30a}$$

and

$$b_1 = 2 + 2\sqrt{1.1/35(35.1)} = 2.05985\,. \tag{30b}$$

(Note: A comprehensive set of tables of $\arcsin x$ is found in [8].) Table A1 of [5] (with $n = 35$, $\alpha = 0.10$ in the notation of that table) yields $\Lambda = 3.5351$. (From the table in section 6.2.1 we see that (19) is approximately 0.501, and we would expect the approximation to be a good one.) Using (28) we compute N^* as the smallest integer equal to or greater than $(3.5351)^2(2.05985)/2(0.0164885)^2 = 47{,}341.7$. Thus in order to guarantee the requirement, one must take at least

47,342 observations, i.e., 47,342 spins of the wheel. (Of course, the last digits in N are not accurate, but they were retained to indicate the method.)

For illustrative purposes we have computed the following table which gives N^* for selected values of k and P^* when $\theta^* = 1.1$ is specified.

P^*	k			
	2	3	6	36
0.75	201	669	2,475	30,775
0.90	724	1,618	4,698	47,342
0.95	1,193	2,389	6,383	59,080
0.99	2,385	4,255	10,303	84,952

9. Generalizations. Thus far in this paper we have considered only Goal 1: "To select the event associated with $p_{[k]}$." However, the same approach can be used in connection with different goals, or more general goals. For example, we might consider Goal 2: "To select the event associated with $p_{[1]}$," or Goal 3: "To select the events associated with $p_{[k-t+2]}, \cdots, p_{[k]}$ *without* regard to order" for any $1 \leqq t \leqq k - 1$. Clearly, Goals 1 and 2 are special cases of Goal 3 since the selection of the $k - 1$ largest is equivalent to the selection of the one smallest. Table A1 of [5], Table 1 of [1], and [9] all provide the constants necessary to deal with Goal 2, but only the two latter tables provide the constants necessary to deal with Goal 3.

10. Acknowledgment. The authors are indebted to Mr. Richard Lesser, Director of the Cornell Computing Center, who supervised the computation of the entries in the various tables.

REFERENCES

[1] R. E. BECHHOFER, "A single-sample multiple-decision procedure for ranking means of normal populations with known variances," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 16–39.

[2] R. E. BECHHOFER AND M. SOBEL, "A single-sample multiple-decision procedure for ranking variances of normal populations," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 273–289.

[3] R. E. BECHHOFER AND M. SOBEL, "A sequential multiple-decision procedure for selecting the multinomial event with the largest probability (preliminary report)," Abstract, *Ann. Math. Stat.*, Vol. 27 (1956), p. 861.

[4] R. E. BECHHOFER AND M. SOBEL, "Non-parametric multiple-decision procedures for selecting that one of k populations which has the highest probability of yielding the largest observation (preliminary report)," Abstract, *Ann. Math. Stat.*, Vol. 29 (1958), p. 325.

[5] S. S. GUPTA, "On a decision rule for a problem in ranking means," *Institute of Statistics Mimeograph Series No. 150*, University of North Carolina, May 1956.

[6] M. J. HUYETT AND M. SOBEL, "Selecting the best one of several binomial populations," *The Bell System Technical Journal*, Vol. 36 (1957), pp. 537–576.

[7] H. KESTEN AND N. MORSE, "A property of the multinomial distribution," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 120–127.

[8] Mathematical Tables Project, U. S. National Bureau of Standards, *Table of Arcsin x*, Columbia University Press, New York, 1945.

[9] D. TEICHROW, *Probabilities associated with order statistics in samples from two normal populations with equal variance*, Chemical Corps Engineering Agency, Army Chemical Center, Maryland, December 1955.

A PROPERTY OF THE MULTINOMIAL DISTRIBUTION

By Harry Kesten[1] and Norman Morse[2]

Cornell University

1. Introduction. The purpose of this paper is to prove a property of the multinomial distribution which is fundamental to the choice of sample size for the selection procedure described in the preceding paper [1] of this issue. As in [1], we let $p_{[1]} \leqq p_{[2]} \leqq \cdots \leqq p_{[k]}$ denote the ranked multinomial probabilities and let $\phi_k = \phi(p_{[1]}, \cdots, p_{[k]})$ be the probability of a correct selection when the selection procedure in section 4 of [1] is used. We wish to prove that for any integers $N \geqq 1$, $k \geqq 2$ and for any number $1 < \theta^* < \infty$, ϕ_k is minimized among all configurations with $p_{[k]} \geqq \theta^* p_{[i]}$ $(i = 1, \cdots, k-1)$ by the configuration $p_{[1]} = p_{[2]} = \cdots = p_{[k-1]} = p_{[k]}/\theta^* = 1/(\theta^* + k - 1)$. This configuration is called "least favorable" because of this property.

The theorem on least favorable configurations, proved below in section 3, merely assembles the results of the preceding lemmas in a rather obvious way. The main ideas of the result are contained in the two lemmas of section 2. Note that Lemma 1, proved below for $k \geqq 3$, is not needed to prove the theorem for the case $k = 2$.

We should like to thank Dr. Milton Sobel for his valuable comments and suggestions.

2. The lemmas. It was found convenient to deviate in the following respect from the notation used in [1]. Let p_k denote the *largest* of the k probabilities, but let the other $(k-1)$ probabilities $p_1, \cdots, p_{k-1}$ be unranked among themselves. Let E_i be the event associated with p_i $(i = 1, \cdots, k)$, and let y_{iN} be the number of occurrences of event E_i after N observations $(i = 1, \cdots, k)$; of course we have $0 \leqq y_{iN} \leqq N$ $(i = 1, \cdots, k)$ and $\sum_{i=1}^{k} y_{iN} = N$. (For notational convenience we use the same symbol for a chance variable and its observed value.) In correspondence to the notation just given, the probability ϕ_k of correctly choosing event E_k is in this paper a function of the p_i's as defined above rather than a function of the $p_{[i]}$'s as it is in [1]. It is sufficient to restrict our attention to configurations with $p_k \geqq \theta^* \max(p_1, \cdots, p_{k-1})$.

Suppose $k \geqq 3$ and any $k-2$ of the p_i's *including* p_k, are held fixed. For notational convenience, we take the two unfixed probabilities to be p_1 and p_2, and we arbitrarily call p_1 the one which is equal to or greater than the other, so that we have $p_1 \geqq p_2$. Note that since the sum $(p_1 + p_2)$ is fixed, ϕ_k can be

Received February 5, 1958.

[1] Research supported by the Office of Naval Research.

[2] Research supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 18(600)-331. Reproduction in whole or in part is permitted for any purpose of the United States Government.

regarded as a function of p_1 only, say $\psi_1(p_1)$, with $(p_1 + p_2)/2 \leqq p_1 \leqq (p_1 + p_2)$.

LEMMA 1. *Let $k \geqq 3$, and let $(p_1 + p_2), p_3, \cdots, p_k$ be fixed as above. Then $\psi_1(p_1)$ is a non-increasing function of p_1, where p_1 lies in the interval $(p_1 + p_2)/2 \leqq p_1 \leqq (p_1 + p_2)$.* [Hence ϕ_k is minimized over all vectors with fixed $p_3, \cdots, p_k$ and such that $p_k \geqq \theta^* \max(p_1, \cdots, p_{k-1})$ by taking $p_1 = \min(p_1 + p_2, p_k/\theta^*)$.]

PROOF OF LEMMA 1. ϕ_k can be rewritten

$$\phi_k = \sum \frac{1}{s} \frac{N!}{w_N!\, y_{2N}!, \cdots, y_{kN}!} q^{w_N} p_3^{y_{2N}}, \cdots, p_k^{y_{kN}} \cdot \sum_{y_{1N}=w_N-y_{kN}}^{y_{kN}} \binom{w_N}{y_{1N}} r^{y_{1N}}(1-r)^{w_N - y_{1N}} S(y_{1N}, y_{2N}, y_{kN}) \tag{1}$$

where $w_N = y_{1N} + y_{2N}$, $q = (p_1 + p_2)$, $r = p_1/q$; the outer summation is over all $(k-1)$-vectors $(w_N, y_{2N}, \cdots, y_{kN})$ such that $w_N + \sum_{i=3}^{k} y_{iN} = N$ and $y_{kN} \geqq y_{iN}$ $(i = 3, \cdots, k-1)$; $(s-1)$ is the number of y_{iN}'s $(i = 3, \cdots, k-1)$ which equal y_{kN}; and $S(y_{1N}, y_{2N}, y_{kN})$ is defined as follows:

$$\begin{aligned} S(y_{1N}, y_{2N}, y_{kN}) &= 1 && \text{if } \max(y_{1N}, y_{2N}) < y_{kN} \\ &= s/(s+1) && \text{if } \max(y_{1N}, y_{2N}) = y_{kN} \text{ and } y_{1N} \neq y_{2N} \\ &= s/(s+2) && \text{if } y_{1N} = y_{2N} = y_{kN} \end{aligned} \tag{2}$$

The inner summation in (1) insures that $y_{kN} \geqq y_{iN}$ for $i = 1, 2$. Note that s is actually a function of $(y_{2N}, \cdots, y_{kN})$.

To prove Lemma 1 it is sufficient to prove the monotonicity property for the inside sum in (1) with w_N and y_{kN} fixed, for then ϕ_k will be termwise monotonic. Using (2), it is easy to see that for w_N and y_{kN} fixed the inside sum in (1) is equal to either

$$\Pr\{w_N - y_{kN} < y_{1N} < y_{kN}\} + \frac{s}{s+1}\Pr\{y_{1N} = w_N - y_{kN}\} + \frac{s}{s+1}\Pr\{y_{1N} = y_{kN}\} \tag{3}$$

for the case $w_N < 2y_{kN}$, or

$$\frac{s}{s+2}\Pr\{y_{1N} = y_{2N} = y_{kN}\} \tag{4}$$

for the case $w_N = 2y_{kN}$. Since the inner sum is empty for $w_N > 2y_{kN}$, no other cases have to be considered.

Now for $w_N \leqq 2a$, define the function

$$f(a; r) = \sum_{x=w_N-a}^{a} \frac{w_N!}{x!(w_N - x)!} r^x (1-r)^{w_N - x}, \tag{5a}$$

which we rewrite as the difference of incomplete Beta functions:

$$(5b)\quad f(a;r) = \frac{w_N!}{a!(w_N - a - 1)!}\int_r^1 [x^a(1-x)^{w_N-a-1} - x^{w_N-a-1}(1-x)^a]\,dx.$$

For the case $w_N > 2a$, we take $f(a; r)$ to be zero. (Also individual terms in (5a) are zero if $x < 0$ or $x > w_N$.) The proof of Lemma 1 can now be completed by noting that expression (3) equals

$$(6)\qquad \frac{1}{s+1}f(y_{kN} - 1; r) + \frac{s}{s+1}f(y_{kN}; r),$$

and that expression (4) equals

$$(7)\qquad \frac{s}{s+2}f(y_{kN}; r).$$

Since the integrand in (5b) is non-negative for the case $w_N \leqq 2a + 1$ and $\frac{1}{2} \leqq r \leqq 1$, it follows that $f(a; r)$ is a non-increasing function of r. The expressions (6) and (7) are therefore non-increasing functions of r, and this proves the lemma.

Starting with any configuration for which $p_k \geqq \theta^* p_i$ $(i = 1, \cdots, k-1)$, one can, by repeated application of Lemma 1 arrive at a new configuration which has $(k - g)$ of the p_i equal to zero, $(g - 2)$ of the p_i equal to p_k/θ^*, and the one remaining p_i in the closed interval $[0, p_k/\theta^*]$. Moreover, the probability of a correct selection for this configuration is at most as large as that for the original configuration. The purpose of the second lemma is to permit further reduction of the probability of a correct selection by starting with this new configuration and making changes of a second type.

To be more precise, we assume that

$$(8)\qquad \begin{gathered} p_k = \theta^* p_{k-1} = \cdots = \theta^* p_{k-g+2} = \frac{\theta^*(1 - p_{k-g+1})}{\theta^* + g - 2} \\ 0 \leqq p_{k-g+1} \leqq p_{k-g+2} = p_k/\theta^* \\ p_{k-g} = \cdots = p_1 = 0, \end{gathered}$$

where g is an integer $2 \leqq g \leqq k$, $\theta^* > 1$, $p_i \geqq 0$ $(i = 1, \cdots, k)$, and $\sum_{i=1}^k p_i = 1$. Note that under these conditions $0 \leqq p_{k-g+1} \leqq 1/(\theta^* + g - 1)$. Let $p_{k-g+1} = p$ to simplify notation. It follows from (8) that ϕ_k may be regarded as a function of p only, say $\psi_2(p)$.

LEMMA 2. *Under the conditions* (8), $\psi_2(p)$ *is a non-increasing function of* p, *where* p *lies in the interval* $0 \leqq p \leqq 1/(\theta^* + g - 1)$. [Hence ϕ_k is minimized under (8) by the configuration $p_1 = \cdots = p_{k-g} = 0$, $p_{k-g+1} = p_{k-g+2} = \cdots = p_{k-1} = p_k/\theta^* = 1/(\theta^* + g - 1)$.]

PROOF OF LEMMA 2. Consider a multinomial problem in which only the $g - 1$ events $E_{k-g+2}, \cdots, E_k$ are involved; suppose the probabilities of these events are given by $1/(\theta^* + g - 2), \cdots, 1/(\theta^* + g - 2), \theta^*/(\theta^* + g - 2)$, respectively. Let M observations be taken, and consider the quantities $y_{k-g+2,M}, \cdots, y_{kM}$. But now let an integer $0 \leqq c \leqq N$ be given which is to correspond to

the event E_{k-g+1}, and use the selection procedure of [1] to choose among the g events $E_{k-g+1}, \cdots, E_k$. That is, the constant c plays the same role in the selection procedure that the chance quantity $y_{k-g+1,M}$ ordinarily would; otherwise the selection is made in the usual way. Let $Q_M(c)$ be the probability of choosing event E_k when this procedure is used.

ϕ_k can now be rewritten as

$$\phi_k = \psi_2(p) = \sum_{y=0}^{N} \binom{N}{y} p^y (1-p)^{N-y} Q_{N-y}(y). \tag{9}$$

Note that $Q_{N-y}(y)$ is independent of p, so that $\psi_2(p)$ is merely a linear combination of binomial terms. One may rewrite (9) as

$$\psi_2(p) = \sum_{y=0}^{N} [Q_{N-y}(y) - Q_{N-y+1}(y-1)] \sum_{m=y}^{N} \binom{N}{m} p^m (1-p)^{N-m}, \tag{10}$$

where we set $Q_{N+1}(-1) = 0$. Since $\sum_{m=y}^{N} \binom{N}{m} p^m (1-p)^{N-m}$ increases with p for all $0 < y \leqq N$ (i.e., the binomial distribution shifts to the right if the probability of a "success" increases), it clearly suffices for proving the lemma to show that $Q_{N-y}(y)$ is non-increasing in y.

From the definition of $Q_M(c)$, it is clear that

$$Q_{N-g-1}(y+1) \leqq Q_{N-g-1}(y). \tag{11}$$

In order to prove $Q_{N-g-1}(y+1) \leqq Q_{N-y}(y)$, therefore, one needs only the inequality $Q_{N-g-1}(y) \leqq Q_{N-y}(y)$. Hence it is sufficient for proving the lemma to prove the following

Assertion. Let $0 \leqq c \leqq N$ be given; then

$$Q_M(c) \leqq Q_{M+1}(c) \tag{12}$$

for all integers M.

To prove the assertion, note that $Q_M(c)$ is the probability that E_k is chosen after M observations; and that if an $(M+1)$st observation were taken, the vector $(y_{k-g+2,M+1}, \cdots, y_{k,M+1})$ could lead to any one of g possible decisions. Likewise, $Q_{M+1}(c)$ is the probability under the same selection procedure that E_k is chosen after $(M+1)$ observations, and this event might have arisen from any of a number of sequences, some of which would have led to E_k, some of which would have led to $E_{k-1}, \cdots$, some of which would haved led to E_{k-g+1}, had the selection been made after the Mth observation. In short, for fixed c, let $R(i,j) = Pr\{\text{choose } E_i \text{ after } M \text{ observations and choose } E_j \text{ after } (M+1) \text{ observations}\}$, where $i, j = k-g+1, \cdots, k$. Then we have

$$\begin{aligned} Q_{M+1}(c) &= \sum_{j=k-g+1}^{k} R(j,k), \\ Q_M(c) &= \sum_{j=k-g+1}^{k} R(k,j); \end{aligned} \tag{13}$$

also

$$Q_{M+1}(c) - Q_M(c) = \sum_{j=k-g+1}^{k} [R(j,k) - R(k,j)]. \tag{14}$$

Thus it suffices for proving Lemma 2 to show that each of the terms

$$R(j,k) - R(k,j) \geqq 0 \qquad (j = k-g+1, \cdots, k-1).$$

In the following, the subscript M is dropped from the cumulative sums y_{iM}; it is understood that the symbols y_i $(i = k-g+1, \cdots, k)$ stand for cumulative sums after M observations. Write $m = \max(c, y_{k-g+2}, \cdots, y_k)$. m is said to occur s times if exactly s of the numbers $c, y_{k-g+2}, \cdots, y_k$ are equal to m and the other y_i's are less than m. The detailed expressions below for $R(j,k)$ and $R(k,j)$ apply only for $j = k-g+2, \cdots, k-1$. It is, of course, necessary to show $R(k-g+1,k) - R(k,k-g+1) \geqq 0$ as well, however we do not write out the expressions here. The argument for this case is much the same as is set out below for the other j, and is in fact somewhat simpler. In the selection procedure which defines $Q_{M+1}(c)$, the only way one can choose E_{k-g+1} after having chosen E_k on the previous observation is through the randomization process. Thus the last two parts of $R(k,k-g+1)$ become zero. This also has the effect of strengthening the desired inequality.

One has, then, for $j = k-g+2, \cdots, k-1$,

$$\begin{aligned} R(j,k) = \sum_{s=1}^{g} \Bigg[& \frac{1}{s^2} \Pr\{y_i = y_k = m\,;\, m \text{ occurs } s \text{ times}; y_i < m-1 \text{ for } y_i \neq m; E_i \text{ occurs at the } (M+1)\text{st observation, for some } i \text{ such that } y_i \neq m\} \\ & + \frac{1}{s(s+1)} \Pr\{y_j = y_k = m; m \text{ occurs } s \text{ times}; y_{i_1} = \cdots = y_{i_u} = m-1; y_i < m-1 \text{ if } y_i \leqq m-1 \text{ and } i \neq i_1, \cdots, i_u\,; E_{i_h} \text{ occurs at the } (M+1)\text{st observation for some } h = 1, \cdots, u\} \\ & + \frac{1}{s(s+1)} \Pr\{y_j = m = y_k + 1; m \text{ occurs } s \text{ times}; E_k \text{ occurs at the } (M+1)\text{st observation}\} \\ & + \frac{1}{s} \Pr\{y_j = y_k = m; m \text{ occurs } s \text{ times}; E_k \text{ occurs at the } (M+1)\text{st observation}\} \Bigg] \end{aligned} \tag{15}$$

Similarly,

$$
\begin{aligned}
R(k,j) = \sum_{s=1}^{g} \Bigg[& \frac{1}{s^2} \Pr\{y_j = y_k = m;\ m \text{ occurs } s \text{ times};\ y_i < m-1 \text{ for } y_i \neq m;\ E_i \text{ occurs at the } (M+1)\text{st observation, for some } i \text{ such that } y_i \neq m\} \\
& + \frac{1}{s(s+1)} \Pr\{y_j = y_k = m;\ m \text{ occurs } s \text{ times};\ y_{i_1} = \cdots = y_{i_u} = m-1;\ y_i < m-1 \text{ if } y_i \leqq m-1 \text{ and } i \neq i_1, \cdots, i_u;\ E_{i_h} \text{ occurs at the } (M+1)\text{st observation for some } h = 1, \cdots, u\} \\
& + \frac{1}{s(s+1)} \Pr\{y_k = m = y_j + 1;\ m \text{ occurs } s \text{ times};\ E_j \text{ occurs at the } (M+1) \text{ observation}\} \\
& + \frac{1}{s} \Pr\{y_j = y_k = m;\ m \text{ occurs } s \text{ times};\ E_j \text{ occurs at the } (M+1)\text{st observation}\} \Bigg]
\end{aligned} \tag{16}
$$

Each of the probabilities $R(j, k)$ and $R(k, j)$ has been divided into four parts, and the parts are now considered in turn. The first two parts of (15) are identical with the first two parts of (16), hence these parts contribute nothing to $R(j, k) - R(k, j)$. As for the third parts, a typical term from the third part of (15) may be written

$$
\frac{1}{s(s+1)} \frac{M!}{(m!)^s (m-1)!\, y_{j_1}! \cdots y_{j_{g-s-1}}!} (\theta^* p_1)^{m-1} p_1^{M-m+1} \cdot \theta^* p_1 , \tag{17}
$$

where $p_1 = 1/(\theta^* + g - 2)$. Similarly a corresponding term from the third part of (16) may be found merely by switching the role of k and j; this term may be written

$$
\frac{1}{s(s+1)} \frac{M!}{(m!)^s (m-1)!\, y_{j_1}! \cdots y_{j_{g-s-1}}!} (\theta^* p_1)^{m} p_1^{M-m} \cdot p_1 , \tag{18}
$$

which is equal to (17). Since the terms in the third parts of (15) and (16) can be put into one-one correspondence, with the corresponding terms being equal, the third parts contribute nothing to $R(j, k) - R(k, j)$. One only has to consider the fourth parts of (15) and (16) therefore.

A typical term from the fourth part of (15) may be written

$$
\frac{1}{s} \frac{M!}{(m!)^s y_{j_1}!, \cdots, y_{j_{g-s}}!} (\theta^* p_1)^{m} p_1^{M-m} \cdot \theta^* p_1 , \tag{19}
$$

while the corresponding term from the fourth part of (16) may be written

$$\frac{1}{s}\frac{M!}{(m!)^s y_{j_1}!, \cdots, y_{j_{g-s}}!}(\theta^* p_1)^m p_1^{M-m} \cdot p_1 . \tag{20}$$

Expression (19) is at least as large as expression (20), since $\theta^* > 1$. Thus the fourth part of (15) is at least as large as the fourth part of (16) and we have $R(j, k) - R(k, j) \geqq 0$. The lemma then follows.

3. Proof of the property of the least favorable configuration.

THEOREM. *Let* $1 < \theta^* < \infty$ *be given, and suppose* $p_k \geqq \theta^* p_i$ $(i = 1, \cdots, k-1)$. *Then*

$$\phi(p_1, \cdots, p_k) \geqq \phi\left(\frac{1}{\theta^* + k - 1}, \cdots, \frac{1}{\theta^* + k - 1}, \frac{\theta^*}{\theta^* + k - 1}\right). \tag{21}$$

PROOF OF THEOREM. Let a vector of probabilities $(p_1, \cdots, p_k)$ be given which satisfies the hypotheses. By applying Lemma 1 repeatedly, one can produce a configuration $(p_1', \cdots, p_k')$ such that $\phi(p_1', \cdots, p_k') \leqq \phi(p_1, \cdots, p_k)$, and which has the following properties:

$$\begin{aligned}
&\text{(a)} && p_k' = p_k && \\
&\text{(b)} && p_i' = p_k/\theta^* && \text{for } h-2 \text{ of the } p_i' \\
&\text{(c)} && p_i' = 0 && \text{for } k-h \text{ of the } p_i' \\
&\text{(d)} && 0 \leqq p_i' \leqq p_k/\theta^* && \text{for the one remaining } p_i'
\end{aligned} \tag{22}$$

where h is an integer $2 \leqq h \leqq k$. If $p_i' = p_k/\theta^*$ for the i in (d) of (22), then the p_i' can be renumbered so that the configuration is in the form of (24), and we can proceed immediately to apply Lemma 2 with $g = h + 1$. If that $p_i' < p_k/\theta^*$, however, we can without loss of generality, renumber the p_i' for $(i < k)$ and rewrite (22) as follows:

$$\begin{aligned}
&\text{(a)} && p_k' = p_k \\
&\text{(b)} && p_{k-1}' = p_{k-2}' = \cdots = p_{k-h+2}' = p_k/\theta^* \\
&\text{(c)} && p_{k-h+1}' = 1 - \sum_{i=k-h+2}^{k} p_i' < p_k/\theta^* \\
&\text{(d)} && p_{k-h}' = p_{k-h-1}' = \cdots = p_1' = 0.
\end{aligned} \tag{23}$$

One can now apply Lemma 2 to the p_i' by substituting h for the g of that lemma. The result is a new configuration $(p_1'', \cdots, p_k'')$ which yields a probability of correct selection at most as large as the one corresponding to $(p_1', \cdots, p_k')$, and which has

$$\begin{aligned}
&\text{(a)} && p_k'' = \theta^* p_{k-1}'' = \cdots = \theta^* p_{k-h+1}'' \\
&\text{(b)} && p_{k-h}'' = p_{k-h-1}'' = \cdots = p_1'' = 0.
\end{aligned} \tag{24}$$

Another application of Lemma 2, with $g = h + 1$, to this new configuration will yield still another configuration whose probability of correct selection is again not larger than the one corresponding to $(p_1'', \cdots, p_k'')$, and for which property (24) holds with h replaced by $h + 1$.

Repeated applications of Lemma 2 lead to a configuration where (24) holds with $h = k$, and this configuration certainly yields $\phi_k \leqq \phi(p_1, \cdots, p_k)$ since the probability of correct selection did not increase at each step. But (24) with $h = k$ is the configuration whose probability of correct selection is given in the right side of (21). Hence the theorem is proved.

REFERENCE

[1] R. E. Bechhofer, S. Elmaghraby, and N. Morse, "A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 102–119.

A CLASSIFICATION PROBLEM INVOLVING MULTINOMIALS[1]

By Oscar Wesler

University of Michigan

1. Introduction. The problem of the k-faced die. There are several distinct classification problems involving multinomials, each known as the problem of the k-faced die. For example, the problem may be to decide whether the die is loaded, and indeed there are several versions of this. One, say, in which the loading is specified; another, in which the unknown loading is estimated from one sample, and then a decision is made as to whether a second sample came from the loaded die or an honest one [1]. Or the problem may be to determine which of the k faces carries a known extra load [2]. Although distinct, all these problems are somewhat related and have certain features more or less in common. The version we shall treat in this paper is still another and a general one of its kind. It will be convenient for our purposes to state it formally as a fixed sample size two-decision statistical problem considered within the framework of composite hypotheses. We shall use the notation and terminology of [3].

Let the space Ω of nature's pure strategies consist of two subspaces Ω_1 and Ω_2, Ω_1 consisting of the $k!$ states got by permuting a known probability distribution $p = (p_1, p_2, \cdots, p_k)$ over the faces $1, 2, \cdots, k$ of a k-faced die, Ω_2 consisting similarly of the $k!$ states arising from a known distribution $q = (q_1, q_2, \cdots, q_k)$. We assume the p_i and q_i strictly positive, and shall further assume, without loss of generality, the vectors p and q written so that $p_1 \geqq p_2 \geqq \cdots \geqq p_k$, and $q_1 \geqq q_2 \geqq \cdots \geqq q_k$. The statistician wishes to make one decision if $\omega \varepsilon \Omega_1$ (the null hypothesis), and another decision if $\omega \varepsilon \Omega_2$ (the alternative hypothesis), the decision to be made on the basis of a sample of N observations $x = (x_1, \cdots, x_N)$, or rather on the basis of the sufficient statistic $r = (r_1, \cdots, r_k)$ representing the number of times each of the k faces appears. Let ϕ be a randomized statistical decision procedure such that if r is observed, the null hypothesis is accepted with probability $\phi(r)$ and the alternative hypothesis is accepted with probability $1 - \phi(r)$. Let $\bar{\alpha}$ and $\bar{\beta}$, the probabilities of the two kinds of errors, be the two functions given by

$$\bar{\alpha}(\omega \mid \phi) = 1 - \sum_r \phi(r) P(r \mid \omega), \qquad \omega \varepsilon \Omega_1$$

$$\bar{\beta}(\omega \mid \phi) = \sum_r \phi(r) P(r \mid \omega), \qquad \omega \varepsilon \Omega_2$$

Received May 12, 1958.

[1] Part of this research was carried out under contract with the Office of Naval Research while the author was at Stanford University, and completed at the University of Michigan with the partial support of the United States Army Signal Air Defense Engineering Agency, Contract No. DA-36-039 SC-64627.

and let

$$\alpha(\phi) = \max_{\omega \in \Omega_1} \bar{\alpha}(\omega \mid \phi),$$

$$\beta(\phi) = \max_{\omega \in \Omega_2} \bar{\beta}(\omega \mid \phi).$$

Then the problem we wish to consider is: Of all procedures ϕ satisfying the condition $\alpha(\phi) = \alpha_0$, to find that procedure ϕ^* which minimizes $\beta(\phi)$.

In this paper a game-theoretic minimax method is used to find an entire class of these desired procedures ϕ^*, optimal in this extended Neyman-Pearson sense. A simplification of the result is then given for the asymptotic case of large N. Finally, for illustrative purposes, a look is taken at the special binomial case of $k = 2$, a case deserving of special attention in its own right. We shall begin with a brief description of the minimax approach, referring the reader to Section 7.7 of [3] for detailed proofs.

2. The minimax method. This method of solving the problem can be understood in terms of a simple geometric picture. Consider the α, β-set in Euclidean 2-space given by $S = \{(\alpha(\phi), \beta(\phi)) : \phi \in \Phi\}$, where Φ is the class of all randomized strategies ϕ. S is a point set in the unit square of the first quadrant and contains the diagonal on which $\alpha + \beta = 1$. Although S is not necessarily convex (unlike the case of a simple hypothesis against a simple alternative) it can be shown that the subset T of S lying on or below the above-mentioned diagonal *is* convex, and that moreover, in the present case, the points on the lower boundary of T, a strictly decreasing convex curve joining the point (0, 1) to the point (1, 0), belong to T. Clearly, it is these points on the lower boundary of T which provide the solution to the statistical problem posed above: Given α_0, we are interested in that procedure ϕ^* which determines the point $(\alpha(\phi^*), \beta(\phi^*))$ on the lower boundary of T whose first coordinate is equal to α_0. Our problem, then, is to characterize the class of test procedures ϕ^* which sweep out the entire lower boundary of T.

Now let G_u be the statistical game that arises when the losses are taken to be 1 for an error of the first kind, a constant $u > 0$ for an error of the second kind, and zero otherwise. It is readily seen from our picture that a minimax solution ϕ_u^* for G_u determines that point on the lower boundary of T for which $\alpha(\phi_u^*) = u\beta(\phi_u^*)$. *It follows at once, therefore, that, by varying u between 0 and ∞, the class of corresponding minimax solutions ϕ_u^* will generate the desired lower boundary of T.*

3. The general solution. The minimax method requires that we solve the statistical game G_u for arbitrary $u > 0$. The problem is clearly invariant under the symmetric group of permutations on k letters. We shall therefore set about finding the class of (unique) invariant minimax procedures which will sweep out the lower boundary of T.[2] For fixed g, with $0 < g < 1$, let λ_g be the a priori probability distribution over Ω which selects Ω_1 with probability g, Ω_2 with prob-

[2] For the specific problem under consideration, it is very easily shown that only invariant procedures need be considered in minimizing $\beta(\phi)$. In more general situations, see [5].

ability $1 - g$, and which is uniform within both Ω_1 and Ω_2, i.e. $\lambda_g(\omega) = g/k!$ for $\omega \varepsilon \Omega_1$ and $\lambda_g(\omega) = (1 - g)/k!$ for $\omega \varepsilon \Omega_2$. Define the distribution function $p(r)$ over the set of all possible outcomes $r = (r_1, \cdots, r_k)$ of the N observations by the formula,

$$p(r) = \frac{1}{k!} \sum_{(i_1, \cdots, i_k)} \frac{N!}{r_1! \cdots r_k!} p_{i_1}^{r_1} p_{i_2}^{r_2} \cdots p_{i_k}^{r_k},$$

where the summation is taken over all permutations $(i_1, \cdots, i_k)$ of the indices $1, \cdots, k$. In other words, $p(r)$ is the average of all the multinomial distributions over the r's arising from the probability vector $p = (p_1, \cdots, p_k)$. $p(r)$ is clearly symmetric in r, that is, $p(r) = p(r_1, r_2, \cdots, r_k) = p(r_{i_1}, r_{i_2}, \cdots, r_{i_k})$. Similarly, define the symmetric distribution $q(r)$ as the average of the multinomial distributions determined by the probability vector $q = (q_1, \cdots q_k)$. The Bayes procedure against λ_g is readily computed in the usual way to be:

$$(1) \qquad \phi_{c,t}(r) = \begin{cases} 1 & \text{if } \dfrac{q(r)}{p(r)} < \dfrac{g}{(1-g)u} = c(g, u) = c \\ 0 & \text{if } \dfrac{q(r)}{p(r)} > c \\ t & \text{if } \dfrac{q(r)}{p(r)} = c \end{cases}$$

where t is any fixed number satisfying $0 \leqq t \leqq 1$. We note that the procedure $\phi_{c,t}$ is a symmetric function of r and depends on g and u only through c. Further, for any choice of t and c, we see that any pair of numbers g_0, u_0 leading to c will yield the same procedure $\phi_{c,t}$, which will be Bayes against λ_{g_0} in the game G_{u_0}. Moreover, c varies between 0 and ∞ as g and u vary respectively between 0 and 1, and 0 and ∞. Varying c between 0 and ∞, and t between 0 and 1, the symmetric partitions of the set of all possible r's into acceptance and rejection regions defined by the inequalities in (1) are clearly such that the acceptance regions vary monotonically from the empty set to the whole space, so that the corresponding α's vary from 1 to 0. By an obvious continuity argument, therefore, given any α_0, with $0 \leqq \alpha_0 \leqq 1$, we can by suitably choosing t find a c such that $\alpha(\phi_{c,t}) = \alpha_0$. Now, consider any one such procedure $\phi_{c,t}$. From the symmetry of $\phi_{c,t}$ it follows that its associated risk function $\rho(\omega, \phi_{c,t})$ is constant over Ω_1 and constant over Ω_2, equal to $\alpha(\phi_{c,t})$ and $u\beta(\phi_{c,t})$ respectively. Selecting u_0 so that $\alpha(\phi_{c,t}) = u_0\beta(\phi_{c,t})$ and then g_0 so that $g_0/(1 - g_0)u_0 = c$, we see that the procedure $\phi_{c,t}$ is Bayes against λ_{g_0} and has a constant risk function over Ω, hence $\phi_{c,t}$ is a minimax procedure for the game G_{u_0}, and determines the point $(\alpha(\phi_{c,t}), \beta(\phi_{c,t}))$ on the lower boundary of T. *By continuity, therefore, the class of invariant procedures $\phi_{c,t}$ with $0 \leqq c < \infty$, $0 \leqq t \leqq 1$, determines the entire lower boundary of T and provides the extended Neyman-Pearson solution to our classification problem.* Finally, since any minimax invariant procedure ϕ^* for a game G_{u_0} must be Bayes against the corresponding λ_{g_0}, and since the only such invari-

ant Bayes procedures are of the form $\phi_{c,t}$, this proves the uniqueness of the class of invariant minimax procedures.

It might perhaps be useful for certain purposes to give the class of invariant partitions or procedures the following somewhat simpler geometric description. Let Ξ be the fundamental probability simplex in k-space, consisting of all vectors $\xi = (\xi_1, \cdots, \xi_k)$ with $\xi_i \geqq 0$, $\sum_{i=1}^{k} \xi_i = 1$, and let $\delta_i = r_i/N$, so that $\delta = (\delta_1, \cdots, \delta_k) \varepsilon \Xi$. By the symmetry of the problem and its solution, we shall assume without loss of generality that $r_1 \geqq \cdots \geqq r_k$, so that the observed probability vector δ as well as the two given vectors p and q all belong to the subsimplex Ξ' of Ξ defined by $\xi_1 \geqq \xi_2 \geqq \cdots \geqq \xi_k$. Then, given α_0, the desired optimal procedure $\phi_{c,t}$ may be characterized in the first place by the simple partition of Ξ' into two regions of δ's (the acceptance region and rejection region, with possible mixing on the boundary) given by the defining inequalities specified in (1); the symmetric images of this partition under all $k!$ permutations then give the corresponding symmetric partition of Ξ which constitutes the whole procedure $\phi_{c,t}$.

4. The case of large N. The kaleidoscopic procedures. For large N, an approximation suitable for most purposes may be given which materially simplifies the class of optimal partitions.

Confining our attention to Ξ', the inequality $q(r)/p(r) < c$ in (1) may be written, on factoring out the largest term from both numerator and denominator, as

$$\left[\left(\frac{q_1}{p_1}\right)^{\delta_1}\left(\frac{q_2}{p_2}\right)^{\delta_2}\cdots\left(\frac{q_k}{p_k}\right)^{\delta_k}\right]^N \frac{1+\Sigma_1}{1+\Sigma_2} < c$$

where, for large N, hence only approximately for large $r_1, r_2, \cdots, r_k$, both Σ_1 and Σ_2 may be taken as zero, or as certain positive integers less than $k!$, depending on the number of equalities existing among the given components q_i of q and among the p_i of p. In any event, taking logarithms, we may then regard the optimal partitions of Ξ' as approximately determined by the one parameter family of hyperplanes

$$\sum_{i=1}^{k} \delta_i \log \frac{q_i}{p_i} = b \qquad (b \text{ an arbitrary constant}). \tag{2}$$

That is, for arbitrary b and t, $0 \leqq t \leqq 1$, the optimal procedure $\phi_{b,t}$ amounts to placing the points δ of Ξ' on the side of the hyperplane where

$$\sum_{i=1}^{k} \delta_i \log \frac{q_i}{p_i} < b$$

into the acceptance region, the points on the other side into the rejection region, those on the hyperplane mixed in the proportion t, $1 - t$, and then taking the symmetric images of this partition to get the symmetric partition of Ξ as a whole. In view of our geometric description, we shall call these approximating $\phi_{b,t}$ the *kaleidoscopic procedures*.

To see at once the reason for this rather striking language, we invite the reader to consider the case of $k = 3$, and to draw a few simple figures, with the points δ or $\xi = (\xi_1, \xi_2, \xi_3)$ of Ξ represented in barycentric coordinates by a point in an equilateral triangle with unit altitude, the distance from the point to the three sides being the values ξ_1, ξ_2 and ξ_3. By selecting different pairs of points for p and q so as to get separating lines with different slopes, and by varying b so as to translate these lines in a parallel fashion, the reason for the term *kaleidoscopic* will appear before his eyes.

In general, by a suitable choice of b and of the mixing factor t, the symmetric or kaleidoscopic test can be made to give us any desired α_0 with approximately the minimum β.

5. The case of $k = 2$: The binomial case. We specialize the previous discussion to $k = 2$ as deserving of special attention, and also to provide a simple illustration of the discussion and results. We refer the reader to problem 7.7.3 of [3] and to pages 75–76 of [4] for a particular example.

We are now in the binomial case and shall speak of coins rather than dice. Two probability distributions, $p = (p_1, p_2)$ and $q = (q_1, q_2)$, are given. Under Ω_1 the coin falls heads with probability either p_1 or $p_2 = 1 - p_1$, under Ω_2 heads with probability either q_1 or $q_2 = 1 - q_1$. Assume things written so that $1 > p_1 > q_1 \geqq 1/2$. Let $r = (r_1, N - r_1)$ represent, respectively, the number of heads and tails appearing among the N tosses of the coin. Then $p(r)$, given by

$$p(r) = \binom{N}{r_1} \frac{p_1^{r_1} p_2^{N-r_1} + p_2^{r_1} p_1^{N-r_1}}{2},$$

and $q(r)$ given similarly, are symmetric bimodal distributions over the number r_1 of heads, $r_1 = 0, 1, \cdots, N$. The likelihood ratio $q(r)/p(r)$ which determines the optimal procedures in (1) may be written

$$\frac{q(r)}{p(r)} = \frac{q_1^{r_1} q_2^{N-r_1} + q_2^{r_1} q_1^{N-r_1}}{p_1^{r_1} p_2^{N-r_1} + p_2^{r_1} p_1^{N-r_1}} = \left(\frac{q_1 q_2}{p_1 p_2}\right)^{N-r_1} \left(\frac{q_1^{2r_1-N} + q_2^{2r_1-N}}{p_1^{2r_1-N} + p_2^{2r_1-N}}\right). \tag{3}$$

Without bothering to convert r to δ, the kaleidoscopic procedures given by (2) are immediately seen to be two-sided symmetric tests of the following form: Take action 1 (decide for p) if $r_1 < j$ or $r_1 > N - j$, action 2 (decide for q) if $j < r_1 < N - j$, and mix actions 1 and 2 in the proportion t, $1 - t$ if $r_1 = j$ or $N - j$, where j is allowed to run over the integers $0, 1, \cdots, [N/2]$ and t is any number satisfying $0 \leqq t \leqq 1$. Denote this two-sided symmetric test by (j, t). Then the kaleidoscopic procedures are given by the set of all pairs (j, t) with $j = 0, 1, \cdots, [N/2]$, $0 \leqq t \leqq 1$. Note that if $p_1 < q_1$ the procedures are of course the same, except that the actions taken are reversed.

The interesting feature here, which we shall prove, is that the optimal procedures of (1) and the (j, t) of (2) are the same, in exact agreement *for every* N, so that there is no question here of approximations or of loss of power in the approximation. To see this, let us agree because of symmetry to consider the likelihood ratio (3) for only those outcomes $r = (r_1, N - r_1)$ for which $2r_1 \geqq N$.

Then the assertion amounts to showing that the inequality $q(r)/p(r) < c$ for any one such r implies the same inequality for all r with a larger r_1. Suppose, therefore, $q(r)/p(r) < c$. Since $q_1q_2 > p_1p_2$, the first factor in the factored form of (3) decreases with increasing r_1. If we can show that the second factor decreases as well with increasing r_1, the assertion will be proved. We require the following simple lemma.

LEMMA: *The function* $y = [x^{m+1} + (1 - x)^{m+1}]/[x^m + (1 - x)^m]$ *considered on the unit interval* $0 < x < 1$, *with* m *an arbitrary non-negative integer, is decreasing for* $0 < x \leqq 1/2$ *and (by symmetry) increasing for* $1/2 \leqq x < 1$.

PROOF: Taking derivatives, a straightforward calculation will show that y' has the same sign as $[x^{2m} - (1 - x)^{2m}] + mx^{m-1}(1 - x)^{m-1}[x^2 - (1 - x)^2]$. Clearly, $y' \leqq 0$ for $0 < x \leqq 1/2$, hence the lemma.

Applying the lemma for successive values of m and multiplying functions together, there follows at once the more general fact that the same decreasing property over $0 < x \leqq 1/2$ and increasing property over $1/2 \leqq x < 1$ is exhibited by the function

$$y = \frac{x^n + (1 - x)^n}{x^m + (1 - x)^m}$$

for arbitrary integers $n > m \geqq 0$. From our assumption of $1/2 \leqq q_1 < p_1 < 1$ there follows the inequality

$$\frac{q_1^n + q_2^n}{q_1^m + q_2^m} < \frac{p_1^n + p_2^n}{p_1^m + p_2^m}$$

or, equivalently, the inequality

$$\frac{q_1^n + q_2^n}{p_1^n + p_2^n} < \frac{q_1^m + q_2^m}{p_1^m + p_2^m}$$

for arbitrary integers $n > m \geqq 0$. But this says that the second factor of (3) decreases with increasing r_1. The asserted agreement between the optimal and kaleidoscopic procedures is thus proved.

It should be pointed out that in practice, given α_0, it becomes a simple matter in this case, using tables of the binomial distribution $P_N(r_1 \mid p_1)$, to calculate the desired test (j, t) from the condition $\alpha(j, t) = \alpha_0$. The minimized β, namely $\beta(j, t)$, is then read off directly from the tables of $P_N(r_1 \mid q_1)$.

REFERENCES

[1] HERMAN CHERNOFF, "A classification problem," *Technical Report 33*, Applied Mathematics and Statistics Laboratory, Stanford University, 1956.

[2] S. G. ALLEN, JR., "A procedure for determining the loaded face of a die," *Technical Report 13*, Applied Mathematics and Statistics Laboratory, Stanford University, 1952.

[3] DAVID BLACKWELL AND M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954, pp. 200–207 in particular.

[4] OSCAR WESLER, *Solutions to Problems in Blackwell and Girshick's Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.

[5] OSCAR WESLER, "Invariance theory and a modified minimax principle," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 1–20.

ON THE LIMITING DISTRIBUTION OF THE NUMBER OF COINCIDENCES CONCERNING TELEPHONE TRAFFIC

By L. Takács

Hungarian Academy of Sciences, Budapest

1. Introduction. Let us consider a telephone exchange. Suppose that the subscribers make calls at the instants $\tau_1, \tau_2, \cdots, \tau_n, \cdots$, where $\tau_n - \tau_{n-1}$ $(n = 1, 2, \cdots; \tau_0 = 0)$ are identically distributed independent positive random variables with the distribution function $F(x)$. Put $\varphi(s) = \int_0^\infty e^{-sx}\, dF(x)$, $\alpha = \int_0^\infty x\, dF(x)$ and $\sigma^2 = \int_0^\infty (x - \alpha)^2\, dF(x)$.

Suppose that there is an infinite number of fully available channels and that each call gives rise to a connection (conversation) on one of the free channels. Denote by χ_n the duration of the holding time beginning in the instant τ_n $(n = 1, 2, \cdots)$. It is assumed that χ_n $(n = 1, 2, \cdots)$ are identically distributed mutually independent positive random variables, which are independent also of the random variables τ_n $(n = 1, 2, \cdots)$. Suppose that $\mathbf{P}\{\chi_n \leqq x\} = 1 - e^{-\mu x}$, if $x \geqq 0$.

We say that the system is in state E_k $(k = 0, 1, 2, \cdots)$ if k channels are busy.

In what follows we shall deal with the determination of the distribution of the number of transitions $E_k \to E_{k+1}$ $(k = 0, 1, 2, \cdots)$ occurring in the time interval $(0, t]$ and the corresponding asymptotic distribution as $t \to \infty$.

The above problem was solved earlier by the author [7] in the particular case when $\{\tau_n\}$ forms a Poisson process with density λ.

2. Notation. Denote by $\eta(t)$ the number of busy channels at the instant t and put

$$\mathbf{P}\{\eta(t) = k\} = P_k(t), \qquad (k = 0, 1, 2, \cdots).$$

Define the r-th binomial moment of $\eta(t)$ as follows:

$$B_r(t) = \sum_{k=r}^{\infty} \binom{k}{r} P_k(t), \qquad (r = 0, 1, 2, \cdots)$$

and put

$$\beta_r(s) = \int_0^\infty e^{-st}\, B_r(t)\, dt, \qquad (\Re(s) > 0).$$

Further denote by $\nu_t^{(k)}$ the number of transitions $E_k \to E_{k+1}$, $(k = 0, 1, 2, \cdots)$, occurring in the time interval $(0, t]$. (We say that a transition $E_{-1} \to E_0$ takes place at $t = 0$.) Denote by $m_k(t)$ the expectation of the random variable $\nu_t^{(k)}$. ($m_{-1}(t) = 1$ if $t \geqq 0$ and $m_{-1}(t) = 0$ if $t < 0$.)

Finally denote by $m(t)$ the expectation of the number of calls taking place in

Received June 20, 1958; revised October 24, 1958.

the time interval $(0, t]$. Then

$$m(t) = \sum_{n=1}^{\infty} F_n(t),$$

where $F_n(t)$ denotes the n-th iterated convolution of $F(t)$ with itself. Clearly,

$$\int_0^{\infty} e^{-st} dm(t) = \frac{\varphi(s)}{1 - \varphi(s)}, \qquad (\Re(s) > 0).$$

3. The solution of the problem. If specifically $\{\tau_n\}$ forms a Poisson process with density λ, then $\{\eta(t)\}$ is a Markov process. In other cases $\{\eta(t)\}$ ceases to be a Markov process, but the instants τ_n always form the Markov points (or regeneration points) of the process. Accordingly, for fixed k ($k = 0, 1, 2, \cdots$), the instants of the successive transitions $E_k \to E_{k+1}$ form a recurrent (or renewal) process.

Denote by $R_k(x)$ the distribution function of the distance between two consecutive transitions $E_k \to E_{k+1}$, and by $R_k^*(x)$ the distribution function of the distance between the first transition $E_k \to E_{k+1}$ and the zero point. Knowing $R_k(x)$ and $R_k^*(x)$, the distribution function of $\nu_t^{(k)}$ can be determined easily; namely, we have

$$\mathbf{P}\{\nu_t^{(k)} > n\} = R_k^* * R_k * \cdots * R_k(t), \tag{1}$$

where the right hand side contains the n-th iterated convolution of $R_k(t)$.

Define

$$\rho_k = \int_0^{\infty} x \, dR_k(x) \tag{2}$$

and

$$\sigma_k^2 = \int_0^{\infty} (x - \rho_k)^2 \, dR_k(x). \tag{3}$$

If $\sigma_k^2 < \infty$, then we have

$$\lim_{t \to \infty} \mathbf{P}\left\{\frac{\nu_t^k - \rho_k t}{\sigma_k t^{1/2}} \leqq x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} \, dy, \tag{4}$$

as it is well known in renewal theory. (Cf. W. Feller [1], W. L. Smith [4], and the author [5].)

If we consider other initial conditions than $\eta(0) = 0$, then we obtain similar results. In particular, the limiting distribution (4) is independent of the initial condition.

Thus, the problem is reduced to the determination of the distribution functions $R_k(x)$ and $R_k^*(x)$. We need some auxiliary theorems, which will be proved below.

4. The Palm functions. Hitherto we have not made any restrictions concerning the servicing of the calls. Now, following C. Palm [3], let us suppose that

the channels are numbered by $1, 2, \cdots, r, \cdots$, and that an incoming call realizes a connection through that idle channel which has the lowest serial number. This assumption does not restrict the generality since $\{\eta(t)\}$ is independent of the system of the handling of traffic. Now denote by $\tau_1^{(r)}, \tau_2^{(r)}, \cdots, \tau_n^{(r)}, \cdots$ the instants of the calls which find all channels busy in the group $(1, 2, \cdots, r)$, leaving the other channels out of consideration. Obviously the time differences $\tau_{n+1}^{(r)} - \tau_n^{(r)}$ $(n = 1, 2, \cdots)$ are identically distributed independent positive random variables. Let us denote by $G_r(x)$ their common distribution function. We shall prove the following

THEOREM 1. *Define*

$$\gamma_r(s) = \int_0^\infty e^{-sx}\, dG_r(x), \qquad (r = 0, 1, 2, \cdots); \tag{5}$$

then we have

$$\gamma_r(s) = \frac{\sum_{j=0}^{r} \binom{r}{j} \prod_{i=0}^{j-1} \frac{1-\varphi(s+i\mu)}{\varphi(s+i\mu)}}{\sum_{j=0}^{r+1} \binom{r+1}{j} \prod_{i=0}^{j-1} \frac{1-\varphi(s+i\mu)}{\varphi(s+i\mu)}}, \tag{6}$$

where the empty product is 1 *and* $G_0(x) \equiv F(x)$.

PROOF. C. Palm [3] has proved that the distribution functions $G_r(x)$ $(r = 1, 2, \cdots)$ satisfy the following system of integral equations:

$$G_r(x) = G_{r-1}(x) - \int_0^x (1 - e^{-\mu y})(1 - G_r(x-y))\, dG_{r-1}(y), \qquad (r = 1, 2, \cdots), \tag{7}$$

where $G_0(x) \equiv F(x)$. This can be proved easily. Let us suppose that $\tau_n^{(r)} = \tau_m^{(r-1)}$ (where $\tau_m^{(0)} = \tau_m$). Then conditionally

$$\mathbf{P}\{\tau_{n+1}^{(r)} - \tau_n^{(r)} \leqq x \mid \tau_{m+1}^{(r-1)} - \tau_m^{(r-1)} = y\} = \begin{cases} e^{-\mu y} + (1 - e^{-\mu y})G_r(x-y), & \text{if } 0 \leqq y \leqq x, \\ 0, & \text{if } y > x, \end{cases}$$

and by the theorem of total probability we have

$$\mathbf{P}\{\tau_{n+1}^{(r)} - \tau_n^{(r)} \leqq x\} = G_r(x) = \int_0^x [e^{-\mu y} + (1 - e^{-\mu y})G_r(x-y)]\, dG_{r-1}(y),$$

which proves (7).

Taking the Laplace-Stieltjes transform of (7) we obtain Palm's recurrence formula,

$$\gamma_r(s) = \frac{\gamma_{r-1}(s+\mu)}{1 - \gamma_{r-1}(s) + \gamma_{r-1}(s+\mu)}, \qquad (r = 1, 2, \cdots), \tag{8}$$

where $\gamma_0(s) = \varphi(s)$.

If we define

$$D_r(s) = \sum_{j=0}^{r} \binom{r}{j} \prod_{i=0}^{j-1} \frac{1-\varphi(s+i\mu)}{\varphi(s+i\mu)}, \qquad (r = 0, 1, 2, \cdots), \tag{9}$$

then it is easy to see that

$$D_{r+1}(s) = D_r(s) + \frac{1-\varphi(s)}{\varphi(s)} D_r(s+\mu), \qquad (r = 0, 1, 2, \cdots). \tag{10}$$

Further, a simple calculation shows that the function,

$$\gamma_r(s) = \frac{D_r(s)}{D_{r+1}(s)}, \qquad (r = 0, 1, 2, \cdots) \tag{11}$$

satisfies (8) and $\gamma_0(s) = \varphi(s)$. This proves (6).

5. The binomial moments $B_r(t)$. We shall prove the following:

THEOREM 2. *The binomial moments $B_r(t)$, $(r = 0, 1, 2, \cdots)$ exist for all $t \geqq 0$, and we have*

$$\beta_r(s) = \int_0^\infty e^{-st} B_r(t)\, dt = \frac{1}{s+r\mu} \prod_{i=0}^{r-1} \frac{\varphi(s+i\mu)}{1-\varphi(s+i\mu)} \tag{12}$$

if $\Re(s) > 0$.

PROOF. Introduce the generating function

$$G(t, z) = \sum_{k=0}^{\infty} P_k(t) z^k. \tag{13}$$

$G(t, z)$ satisfies the following integral equation:

$$G(t, z) = [1 - F(t)] + \int_0^t G(t-x, z)[1 - (1-z)e^{-\mu(t-x)}]\, dF(x). \tag{14}$$

This can be proved as follows. Define

$$f(t, u) = \begin{cases} 1, & \text{if } 0 \leqq t \leqq u, \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$\eta(t) = \sum_{n=1}^{\infty} f(t - \tau_n, \chi_n).$$

Now let us suppose conditionally that $\tau_1 = x$; then

$$\eta(t) = \begin{cases} f(t-x, \chi_1) + \bar{\eta}(t-x), & \text{if } x \leqq t, \\ 0, & \text{if } x > t, \end{cases}$$

where $\bar{\eta}(t-x)$ is independent of $f(t-x, \chi_1)$ and has the same distribution as $\eta(t-x)$. Here the generating function of $f(t-x, \chi_1)$ is $[1 - e^{-\mu(t-x)} + ze^{-\mu(t-x)}]$, if $0 \leqq x \leqq t$, and the generating function of $\bar{\eta}(t-x)$ is $G(t-x, z)$, if $0 \leqq x \leqq t$.

Therefore, applying the theorem of total expectation for $G(t, z) = \mathbf{E}\{z^{\nu(t)}\}$, we obtain

$$G(t, z) = [1 - F(t)] + \int_0^t G(t - x, z)[1 - e^{-\mu(t-x)} + ze^{-\mu(t-x)}]\, dF(x),$$

which proves (14).

I am indebted to R. Syski for calling my attention to the possibility of the above proof. Applying the results of R. Fortet [2] or the author [6], R. Syski showed that $G(t, z)$ satisfies the following integral equation:

$$G(t, z) = 1 - (1 - z)\int_0^t G(t - x, z)e^{-\mu(t-x)}dm(x), \tag{15}$$

where $m(t)$ denotes the expected number of the calls occurring in the time interval $(0, t]$.

Since

$$B_r(t) = \frac{1}{r!}\left(\frac{d^r G(t, z)}{dz^r}\right)_{z=1}, \qquad (r = 0, 1, 2, \cdots), \tag{16}$$

we obtain from (14) that

$$B_r(t) = \int_0^t B_r(t - x)\, dF(x) + \int_0^t B_{r-1}(t - x)e^{-\mu(t-x)}dF(x), \quad (r = 1, 2, \cdots). \tag{17}$$

This is a linear integral equation of the Volterra type for the unknown $B_r(t)$. As is well known, the solution is

$$B_r(t) = \int_0^t B_{r-1}(t - x)e^{-\mu(t-x)}\, dm(x), \qquad (r = 1, 2, \cdots). \tag{18}$$

This can be obtained immediately from Syski's equation (15).

Taking the Laplace transform of (18), we obtain the following functional equation:

$$\beta_r(s) = \frac{\varphi(s)}{1 - \varphi(s)}\,\beta_{r-1}(s + \mu). \tag{19}$$

Since $B_0(t) \equiv 1$, consequently $\beta_0(s) = 1/s$, and applying repeatedly formula (19) we finally obtain

$$\beta_r(s) = \frac{1}{s + r\mu}\prod_{i=0}^{r-1}\frac{\varphi(s + i\mu)}{1 - \varphi(s + i\mu)},$$

as was to be proved.

It is to be remarked that there exists a constant C so that

$$B_r(t) < \frac{C^r}{r!}, \qquad (r = 0, 1, 2, \cdots), \tag{20}$$

for all $t \geqq 0$. This can be proved by virtue of (18)

Remark. Since

$$B_r(t) = \sum_{k=r}^{\infty} \binom{k}{r} P_k(t) \tag{21}$$

and $B_r(t) < C^r / r!$, $(r = 0, 1, 2, \cdots)$, we obtain easily that

$$P_k(t) = \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} B_r(t). \tag{22}$$

Hence, specifically,

$$\int_0^{\infty} e^{-st} P_k(t)\, dt = \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} \frac{1}{s + r\mu} \sum_{i=0}^{r-1} \frac{\varphi(s + i\mu)}{1 - \varphi(s + i\mu)}. \tag{23}$$

6. The transitions $E_k \to E_{k+1}$.

THEOREM 3. *If $m_k(t)$ denotes the expectation of the number of transitions $E_k \to E_{k+1}$ occurring in the time interval $(0, t]$, then we have*

$$\int_0^{\infty} e^{-st}\, dm_k(t) = \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} \prod_{i=0}^{r} \frac{\varphi(s + i\mu)}{1 - \varphi(s + i\mu)}, \quad (k = 0, 1, 2, \cdots). \tag{24}$$

PROOF. Applying the theorem of total probability we can write

$$P_k(t) = \sum_{j=k}^{\infty} \binom{j}{k} \int_0^t e^{-k\mu(t-u)} [1 - e^{-\mu(t-u)}]^{j-k} [1 - F(t - u)]\, dm_{j-1}(u). \tag{25}$$

This follows from the fact that the event that there is a state E_k at the instant t can occur in several mutually exclusive ways: the last transition in the time interval $(0, t]$ is $E_{j-1} \to E_j$ $(j = k, k + 1, \cdots)$ and this transition takes place at the instant $u (0 \leqq u \leqq t)$, and in the time interval $(u, t]$ there does not occur any new call, but $j - k$ conversations terminate.

Hence,

$$B_r(t) = \sum_{k=r}^{\infty} \binom{k}{r} P_k(t) = \sum_{j=r}^{\infty} \binom{j}{r} \int_0^t e^{-r\mu(t-u)} [1 - F(t - u)]\, dm_{j-1}(u), \tag{26}$$

where $\binom{k}{r}\binom{j}{k} = \binom{j}{r}\binom{j - r}{k - r}$ has been used. Forming the Laplace transform of (26), we have

$$\beta_r(s) = \frac{1 - \varphi(s + r\mu)}{s + r\mu} \sum_{j=r}^{\infty} \binom{j}{r} \int_0^{\infty} e^{-st}\, dm_{j-1}(t).$$

Now by the aid of (12) we obtain

$$\sum_{j=r}^{\infty} \binom{j}{r} \int_0^{\infty} e^{-st}\, dm_{j-1}(t) = \frac{1}{1 - \varphi(s + r\mu)} \prod_{i=0}^{r-1} \frac{\varphi(s + i\mu)}{1 - \varphi(s + i\mu)}.$$

Multiplying both sides of this formula by $(-1)^{r-l}\binom{r}{l}$ and summing over $r =$

$l, l+1, \cdots$ we obtain

$$\int_0^\infty e^{-st}\, dm_{l-1}(t)$$

$$= \sum_{r=l}^{\infty} (-1)^{r-l} \binom{r}{l} \left[\prod_{i=0}^{r-1} \frac{\varphi(s+i\mu)}{1-\varphi(s+i\mu)} + \prod_{i=0}^{r} \frac{\varphi(s+i\mu)}{1-\varphi(s+i\mu)} \right]. \tag{27}$$

If we write $l = k+1$, then

$$\int_0^\infty e^{-st}\, dm_k(t) = \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} \prod_{i=0}^{r} \frac{\varphi(s+i\mu)}{1-\varphi(s+i\mu)}, \tag{28}$$

which was to be proved.

7. The distributions $R_k(x)$ and $R_k^*(x)$.

THEOREM 4. *We have*

$$\int_0^\infty e^{-sx}\, dR_k^*(x) = \left[\sum_{j=0}^{k+1} \binom{k+1}{j} \prod_{i=0}^{j-1} \frac{1-\varphi(s+i\mu)}{\varphi(s+i\mu)} \right]^{-1} \tag{29}$$

and

$$\int_0^\infty e^{-sx}\, dR_k(x) = 1 - \left\{ \left[\sum_{j=0}^{k+1} \binom{k+1}{j} \prod_{i=0}^{j-1} \frac{1-\varphi(s+i\mu)}{\varphi(s+i\mu)} \right] \cdot \left[\sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} \prod_{i=0}^{r} \frac{\varphi(s+i\mu)}{1-\varphi(s+i\mu)} \right] \right\}^{-1}, \tag{30}$$

if $\Re(s) \geqq 0$.

PROOF. Denote by $G_0(x), G_1(x), \cdots, G_k(x)$ the distribution functions of the distances between the successive transitions $E_{-1} \to E_0$, $E_0 \to E_1$, $E_1 \to E_2$, $\cdots$, $E_k \to E_{k+1}$, respectively. It is easy to see that the distribution functions $G_r(x)$ $(r = 0, 1, \cdots, k)$ are just Palm's distribution functions defined by (6). Now clearly

$$R_k^*(x) = G_0 * G_1 * \cdots * G_k(x), \tag{31}$$

and thus,

$$\int_0^\infty e^{-sx}\, dR_k^*(x) = \gamma_0(s)\gamma_1(s) \cdots \gamma_k(s), \tag{32}$$

where $\gamma_r(s)$ $(r = 0, 1, 2, \cdots)$ is defined by (6). This proves (29).

On the other hand, if

$$\Psi_k(s) = \int_0^\infty e^{-sx}\, dR_k(x), \tag{33}$$

then we have

$$\int_0^\infty e^{-st}\, dm_k(t) = \frac{\gamma_0(s)\gamma_1(s) \cdots \gamma_k(s)}{1-\Psi_k(s)}. \tag{34}$$

For, as is well known in renewal theory, we have

$$m_k(t) = R_k^*(t) + R_k^* * R_k(t) + R_k^* * R_k * R_k(t) + \cdots .$$

Taking into consideration (6) and (24), we can determine $\Psi_k(s)$ from (34), and thus we get (30).

THEOREM 5. *We have*

$$\rho_k = \frac{\alpha}{\sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} C_r} \tag{35}$$

and

$$\sigma_k^2 = 2\alpha\rho_k \sum_{j=1}^{k+1} \binom{k+1}{j} C_{j-1} - \rho_k^2 \left[1 - \frac{\sigma^2 - \alpha^2}{\alpha^2} \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} C_r - \frac{2}{\alpha} \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} C_r \sum_{i=1}^{r} \frac{\varphi'(i\mu)}{\varphi(i\mu)[1-\varphi(i\mu)]} \right], \tag{36}$$

where

$$C_r = \prod_{i=1}^{r} \frac{1 - \varphi(i\mu)}{\varphi(i\mu)}, \qquad (r = 0, 1, 2, \cdots). \tag{37}$$

PROOF. Since

$$\sum_{j=0}^{k+1} \binom{k+1}{j} \prod_{i=0}^{j-1} \frac{1 - \varphi(s + i\mu)}{\varphi(s + i\mu)} = 1 + s\alpha \sum_{j=1}^{k+1} \binom{k+1}{j} C_{j-1} + o(s)$$

and

$$\sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} \prod_{i=0}^{r} \frac{\varphi(s + i\mu)}{1 - \varphi(s + i\mu)} = \frac{1}{s\alpha} \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} C_r + \frac{\sigma^2 - \alpha^2}{2\alpha^2} \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} C_r + \frac{1}{\alpha} \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} C_r \sum_{i=1}^{r} \frac{\varphi'(i\mu)}{\varphi(i\mu)[1-\varphi(i\mu)]} + o(s),$$

as $s \to 0$, we obtain easily that

$$\Psi_k(s) = 1 - \rho_k s + \frac{\sigma_k^2 + \rho_k^2}{2} s^2 + o(s^2),$$

as $s \to 0$, where ρ_k and σ_k^2 are defined by (35) and (36) respectively. This proves the theorem.

REFERENCES

[1] W. FELLER, "Fluctuation theory of recurrent events," *Trans. Amer. Math. Soc.*, Vol. 67 (1949), pp. 98–119.

[2] R. FORTET, "Random distributions with an application to telephone engineering," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. II, University of California Press, 1956, pp. 81–88.

[3] C. PALM, "Intensitätsschwankungen im Fernsprechverkehr," *Ericsson Technics* No. 44 (1943), pp. 1–189.

[4] W. L. SMITH, "Regenerative stochastic processes," *Proc. Roy. Soc. Edinburgh. Sec. A.*, 232 (1955), pp. 6–31.

[5] L. TAKÁCS, "On a probability problem arising in the theory of counters," *Proc. Cambridge Philos. Soc.*, Vol. 52 (1956), pp. 488–498.

[6] L. TAKÁCS, "On secondary stochastic processes generated by recurrent processes," *Acta Mathematica Academiae Scientiarum Hungaricae*, Vol. 7 (1956), pp. 17–29.

[7] L. TAKÁCS, "On a coincidence problem concerning telephone-traffic," *Acta Mathematica Academiae Scientiarum Hungaricae*, Vol. 9 (1958), pp. 45–81.

ON THE INTEGRODIFFERENTIAL EQUATION OF TAKÁCS. II[1]

By Edgar Reich

University of Minnesota

1. Introduction. This paper continues (Cf [1]) the study of the properties of the function, $F(t, 0)$, $0 \leqq F(t, 0) = F(t) \leqq 1$, where $F(t, x) = \Pr\{\eta(t) \leqq x\}$, $t \geqq 0$, $x \geqq 0$, satisfies[2]

$$\frac{\partial F(t,x)}{\partial t} = \frac{\partial F(t,x)}{\partial x} - \lambda(t)F(t,x) + \lambda(t)\int_{0-}^{x} H(x-y)\, d_y F(t,y)\,. \tag{1.1}$$

The functions $\Phi(s) = \int_0^\infty e^{-sx}\, dF(0,x)$, $H(x) = \int_0^x h(\xi)\, d\xi$,

$$(h(\xi) \geqq 0,\ \textstyle\int_0^\infty h(\xi)\, d\xi = 1),$$

and $\lambda(t) = \Lambda'(t) \geqq 0$ are given. It is assumed that there exists a $c > 0$ such that $e^{-cx}\, h(x)\ \varepsilon\ L^2(0, \infty)$. The moment $\int_0^\infty x^k h(x)\, dx$, if it exists, is denoted by μ_k. We put $\psi(s) = \int_0^\infty h(x)e^{-sx}\, dx$. Furthermore, we suppose that $\int_0^T [\lambda(t)]^2\, dt$ exists as a possibly improper Riemann integral for all $T > 0$.

The stochastic process $\eta(t)$ represents the waiting time of a customer arriving at time t in a queue with Poisson arrivals of variable density $\lambda(t)$, with $H(x)$ the distribution of service times. $F(t)$ is the probability that the counter is unoccupied at time t. Our present purpose is to study the behavior of $F(t)$ for large t, especially under conditions that turn out to guarantee that $F(t)$ does not approach zero. Previous knowledge ([2], [3], [4]) in this direction appears to be restricted essentially to the case $\lambda(t) = \text{const.}$

The following was proved in [1], although it does not appear as an explicit statement there.

Theorem 1. *There is only one distribution-solution, $F(t, x)$, of (1.1). Moreover, $F(t)$ is the unique continuous solution of the Volterra equation of the first kind*

$$\int_0^t G(t,u)F(u)\, du = g(t), \qquad (\textit{almost all } t \geqq 0)\,, \tag{1.2}$$

where

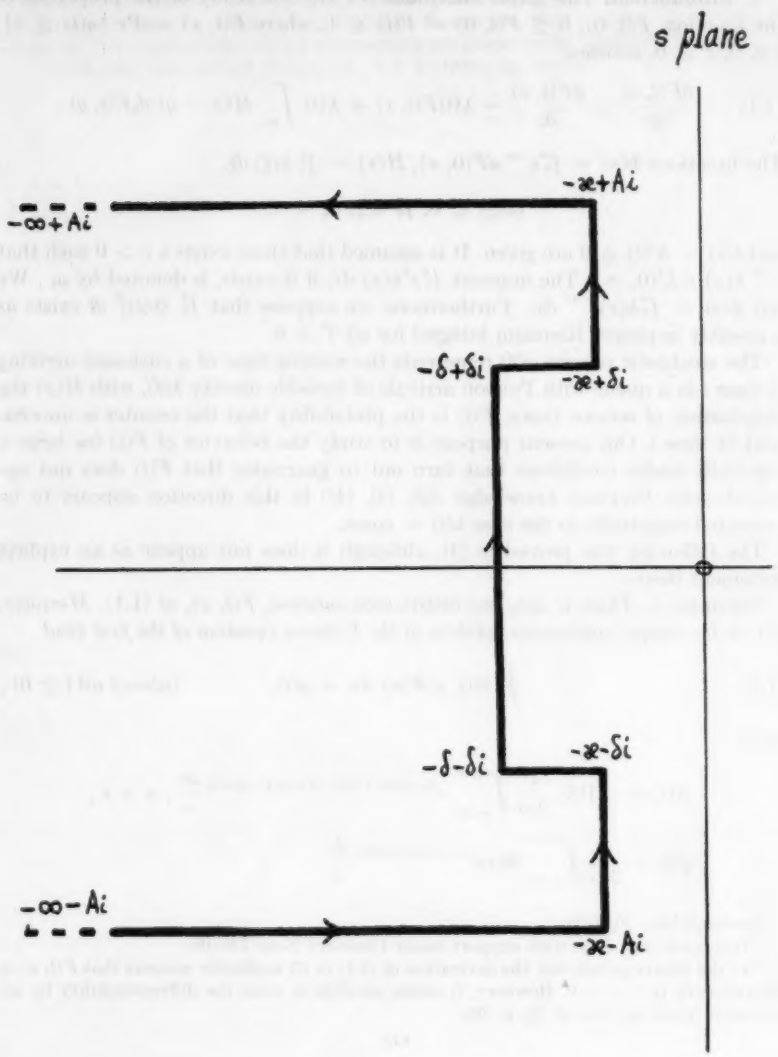
$$\begin{cases} G(t,u) = \text{P.V.}\,\dfrac{1}{2\pi i}\displaystyle\int_{x-i\infty}^{x+i\infty} e^{(t-u)s-[\Lambda(t)-\Lambda(u)][1-\psi(s)]}\,\dfrac{ds}{s}\,,\ x > c\,, \\ g(t) = \dfrac{1}{2\pi i}\displaystyle\int_{x-i\infty}^{x+i\infty} \Phi(s)e^{ts-\Lambda(t)[1-\psi(s)]}\,\dfrac{ds}{s^2}\,. \end{cases}$$

Received May 20, 1958.

[1] This work was done with support under Contract Nonr-710(16).

[2] As the referee points out, the derivation of (1.1) in [2] implicitly *assumes* that $F(t, x)$ is differentiable in x, $x > 0$. However, it seems possible to *prove* the differentiability by an argument based on that of [2], p. 108.

Our approach to the asymptotic behavior of $F(t)$ will be through (1.2). This will, of course, not be as simple as in the case, $\lambda(t) = \text{const}$, when (1.2) can be solved by Laplace transforms. The main results shall be based on the restrictive assumption that $\psi(s)$ is regular at infinity, and in a neighborhood of the imaginary axis. The class of such $\psi(s)$ is, however, still large enough to include as a proper subclass the important class of rational $\psi(s)$.



2. Some Abelian lemmas.

LEMMA 2.1. *Suppose $\psi(s)$ is regular at infinity and in a neighborhood of the imaginary axis, and $\beta\mu_1/\alpha \leqq \rho < 1$. Then there exist functions $k_j(\rho) > 0, j = 1, 2$, such that*

$$\left| \text{P.V.} \frac{1}{2\pi i} \int_{x-i\infty}^{x+i\infty} e^{\alpha s - \beta[1-\psi(s)]} \frac{ds}{s} - 1 \right| \leqq k_1(\rho) e^{-k_2(\rho)\alpha}, \alpha \geqq 0, \beta \geqq 0.$$

PROOF. If $A > 0$ and $\delta > 0$ are chosen so that ψ is regular for $|\Im s| \geqq A$ and $\Re s \geqq -\delta$, we can write the quantity inside the absolute value signs as $H(\alpha, \beta) = (1/2\pi i) \int_C$, where $C = C(A, \delta, \varkappa)$ is the contour shown in the figure.

Since $\psi(\infty) = 0$ we may choose A sufficiently large so that $|\psi(x \pm iA)| < 1$. The function $\psi(s)$ is of the form $\psi(s) = 1 - \mu_1 s + \mu_2/2s^2 + O(s^3)$, $\mu_i > 0$. Recalling that $\mu_2 > 0$, let us choose δ sufficiently small so that

$$\Im \frac{\psi - 1 + \mu_1 s}{s} > 0, \tag{2.1}$$

and < 0, on the segments $[-\delta + \delta i, \delta i]$, and $[-\delta - \delta i, -\delta i]$, respectively. Furthermore, we choose $\delta = \delta(\rho)$ sufficiently small so that, also,

$$\max_{|s| \leqq \sqrt{2}\delta} \left| \frac{\psi - 1 + \mu_1 s}{\mu_1 s} \right| \leqq \frac{1}{\sqrt{2}} (\rho^{-1/2} - 1).$$

Finally, we select the number $x = x(\rho)$, $0 < x < \delta$, so that

$$\Re\psi(-x + iy) < 1 \qquad \text{when } \delta \leqq |y| \leqq A.$$

(This is possible since $\Re\psi(iy) = \int_0^\infty \cos(y\xi) h(\xi)\, d\xi < 1$ when $|y| > 0$.) In the proof of the lemma we may assume that $\alpha \geqq 1$, because in the opposite case both α and β are bounded from above, and the truth of the lemma follows from the fact that, because of the continuity of $H(\alpha, \beta)$, at most an adjustment of $k_i(\rho)$ is necessary.

On the segment $[-\delta - \delta i, -\delta + \delta i]$,

$$\Re\{\alpha s - \beta[1 - \psi(s)]\} = \Re\left\{ \left[\alpha - \beta\mu_1 \left(1 - \frac{\psi - 1 + \mu_1 s}{\mu_1 s} \right) \right] s \right\}$$

$$\leqq (\alpha - \beta\mu_1)(-\delta) + \beta\mu_1 |s| \left| \frac{\psi - 1 + \mu_1 s}{\mu_1 s} \right| \leqq (\alpha - \beta\mu_1)(-\delta)$$

$$+ \beta\mu_1 \sqrt{2}\delta \frac{1}{\sqrt{2}} (\rho^{-1/2} - 1) \leqq -\alpha\delta(1 - \rho^{1/2}).$$

Thus

$$\left| \int_{-\delta-\delta i}^{-\delta+\delta i} \right| \leqq \frac{2\delta}{\delta} e^{-\alpha\delta(1-\rho^{1/2})} = c_1(\rho) \exp[-c_2(\rho)\alpha], c_2(\rho) > 0.$$

On the segments $[-\delta \pm \delta i, -x \pm \delta i]$ we find, by (2.1), that

$$\Re\{\alpha s - \beta[1 - \psi(s)]\} = (\alpha - \beta\mu_1)\Re(s) + \beta\mu_1 \left[\Re(s)\Re\left(\frac{\psi - 1 + \mu_1 s}{\mu_1 s} \right) \right.$$

$$\left. \mp \delta\Im\left(\frac{\psi - 1 + \mu_1 s}{\mu_1 s} \right) \right] \leqq \alpha \left\{ 1 - \frac{\beta\mu_1}{\alpha} \left[1 - \Re\left(\frac{\psi - 1 + \mu_1 s}{\mu_1 s} \right) \right] \right\} \Re s.$$

Since $\Re s < 0$, the above

$$\leqq \alpha\left\{1 - \frac{\beta\mu_1}{\alpha}\left[1 + \left|\frac{\psi - 1 + \mu_1 s}{\mu_1 s}\right|\right]\right\} \Re s \leqq \alpha\{1 - \rho[1 + \rho^{-1/2} - 1]\} \Re s \leqq$$

$$-\alpha(1 - \rho^{1/2})x.$$

Hence

$$\left|\int_{-x-\delta i}^{-\delta-\delta i} + \int_{-\delta+\delta i}^{-x+\delta i}\right| \leqq \frac{2(\delta - x)}{\delta} e^{-\alpha(1-\rho^{1/2})\alpha} = \mathfrak{C}_3(\rho)e^{-\mathfrak{C}_4(\rho)\alpha}, \qquad \mathfrak{C}_4(\rho) > 0.$$

In view of the choice of x,

$$\left|\int_{-\alpha-\Delta i}^{-\alpha-\delta i} + \int_{-\alpha+\delta i}^{-\alpha+\Delta i}\right| \leqq \frac{2A}{\delta} e^{-\alpha\alpha} = \mathfrak{C}_5(\rho)e^{-\mathfrak{C}_6(\rho)\alpha}, \quad \mathfrak{C}_6(\rho) > 0.$$

Finally,

$$\left|\int_{-\infty-\Delta i}^{-\alpha-\Delta i} + \int_{-\alpha+\Delta i}^{-\infty+\Delta i}\right| \leqq \frac{2}{A}\int_{\alpha}^{\infty} e^{-\alpha x}dx = \mathfrak{C}_7 e^{-\mathfrak{C}_8(\rho)\alpha}, \mathfrak{C}_8(\rho) > 0, \text{ if } \alpha \geqq 1.$$

Putting all the above inequalities together, we obtain the lemma.

COROLLARY. *Suppose $\psi(s)$ is regular at infinity and in a neighborhood of the imaginary axis, and*

$$\limsup_{t-u\to\infty}\left[\mu_1\frac{\Lambda(t) - \Lambda(u)}{t - u}\right] < 1.$$

Then

$$\int_0^t G(t,u)F(u)\,du = \int_0^t F(u)\,du + O(1), \qquad \text{as } t \to \infty. \tag{2.2}$$

PROOF. There exists a $T > 0$ such that

$$\mu_1\frac{\Lambda(t) - \Lambda(u)}{t - u} \leqq \rho < 1 \text{ whenever } t - u \geqq T.$$

If $|G(t,u)| \leqq M$ for $|t - u| \leqq T$ then, for $t > T$,

$$\left|\int_0^t G(t,u)F(u)\,du - \int_0^t F(u)\,du\right| \leqq \left|\int_0^{t-T}[G(t,u) - 1]F(u)\,du\right| + (M+1)T$$

$$\leqq \int_0^{t-T} k_1(\rho)e^{-k_2(\rho)(t-u)}du + (M+1)T \leqq \frac{k_1(\rho)}{k_2(\rho)} + (M+1)T = O(1).$$

LEMMA 2.2. *Suppose $\psi(s) = 1 - \mu_1 s + o(s)$, as $s \to 0$, uniformly with respect to* $\arg s$, $|\arg s| \leqq \pi/2$, *and $\mu_1\Lambda(t)/t \leqq \rho < 1$. Then $g(t) = t - \mu_1\Lambda(t) + o(t)$, as $t \to \infty$.*

PROOF. Let C_r denote a contour along the imaginary axis, with a semicircular detour of radius r, center 0, to the right. Since $|\Phi(s)| \leqq |$ for $\Re s \geqq 0$, we have,

according to a known Abelian theorem ([5], pp. 494–5),

$$g_1(t) = \frac{1}{2\pi i}\int_{C_r} \Phi(s)e^{[t-\mu_1\Lambda(t)]s}\frac{ds}{s^2} = t - \mu_1\Lambda(t) + o(t), \qquad \text{as } t \to \infty .$$

It therefore suffices to show that

$$D(t) = g(t) - g_1(t) = \frac{1}{2\pi i}\int_{C_r} \Phi(s)e^{(t-\mu_1\Lambda)s}[e^{(\psi-1+\mu_1 s)\Lambda} - 1]\frac{ds}{s^2} = o(t), \text{ as } t \to \infty.$$

Choose $\epsilon > 0$. Then, if $|s| \leqq \delta(\epsilon)$, $|\psi - 1 + \mu_1 s| \leqq \epsilon|s|$, and therefore

$$|e^{(\psi-1+\mu_1 s)\Lambda} - 1| \leqq \Lambda\epsilon|s|e^{\Lambda\epsilon|s|}.$$

Consider points $\pm iA$, $0 < r < A \leqq \delta(\epsilon)$ on C_r. Evidently,

$$\left|\int_{-iA}^{iA}\right| \leqq e^{(t-\mu_1\Lambda)r}\Lambda\epsilon e^{\Lambda\epsilon A}\int_{-iA}^{iA}\frac{|ds|}{|s|} = \Lambda\epsilon e^{(t-\mu_1\Lambda)r+\Lambda\epsilon A}\left(\pi + 2\log\frac{A}{r}\right).$$

Also,

$$\left|\int_{-i\infty}^{-iA} + \int_{iA}^{i\infty}\right| \leqq 2\int_A^\infty |e^{ts-(1-\psi)\Lambda} - e^{(t-\mu_1\Lambda)s}|\frac{dy}{y^2} \leqq 4\int_A^\infty \frac{dy}{y^2} = \frac{4}{A}.$$

Hence,

$$D(t) \leqq \frac{4}{A} + \Lambda\epsilon e^{(t-\mu_1\Lambda)r+\Lambda\epsilon A}\left(\pi + 2\log\frac{A}{r}\right).$$

Putting $r = t^{-1}$, $A = \epsilon^{-1}t^{-1}$, where $t \geqq T(\epsilon) = [\epsilon\delta(\epsilon)]^{-1}$, we find that

$$D(t) \leqq Kt\epsilon\log(1/\epsilon) \text{ for } t \geqq T(\epsilon).$$

We shall state the following without proof.

LEMMA 2.3. *If $\psi(s)$ and $\Phi(s)$ are regular in a neighborhood of $s = 0$, and if $\mu_1\Lambda(t)/t \leqq \rho < 1$ then $g(t) = t - \mu_1\Lambda(t) + O(1)$, as $t \to \infty$.*

3. Asymptotic average of $F(t)$. In the special case $\lambda(t) = \lambda = \text{const}$ it is known [4] that, for the distribution solution $F(t, x)$ of (1.1), $\mu_1\lambda < 1$ implies $\lim_{t\to\infty}F(t, 0) = 1 - \mu_1\lambda$. In view of our above results we are able to make the following statements in the general case.

THEOREM 2. *Suppose $\psi(s)$ is regular at infinity and in a neighborhood of the imaginary axis, and*

$$\limsup_{t-u\to\infty} \mu_1\frac{\Lambda(t) - \Lambda(u)}{t - u} < 1.$$

Then, if $F(t, x)$ is the distribution solution of (1.1),

$$\frac{1}{T}\int_0^T F(t, 0)\,dt = 1 - \mu_1\frac{\Lambda(T)}{T} + o(1), \text{ as } T \to \infty .$$

THEOREM 3. *If we also assume that $\Phi(s)$ is regular in a neighborhood of $s = 0$ then "$o(1)$" in the conclusion of Theorem 2 can be replaced by "$O(T^{-1})$."*

REFERENCES

[1] E. Reich, "On the Integrodifferential Equation of Takács I", *Ann. Math. Stat.*, Vol. 29 (1958), pp. 563–570.

[2] L. Takács, "Investigation of waiting time problems by reduction to Markov processes", *Acta Math. Hung.*, Vol. 6 (1955), pp. 101–129.

[3] D. G. Kendall, "Some problems in the theory of queues," *J. Roy. Stat. Soc.* (B), Vol. 13 (1951), pp. 151–173.

[4] V. E. Beneš, "On queues with Poisson arrivals," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 670–677.

[5] G. Doetsch, *Handbuch der Laplace-Transformation I*, Birkhäuser Verlag, 1950.

SYMMETRIZABLE MARKOV MATRICES

By H. P. Kramer

Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

Introduction. Suppose that the evolution of the state probabilities $p_i(t)$ of a Markov process is governed by the system of differential equations

$$\frac{dp_i}{dt} = \sum_{j=1}^{N} Q_{ij}p_j(t), \qquad i = 1, \cdots, N, \tag{1}$$

where Q_{ij} represents the transition rate from state S_j to state S_i ([1], p. 235). In many applications one is interested primarily in knowing the equilibrium state probabilities π_i defined by $\pi_i = \lim_{t\to\infty} p_i(t)$, which, if they exist, can be obtained by solving the system of homogeneous linear equations

$$\sum_{j=1}^{N} Q_{ij}\pi_j = 0, \qquad i = 1, 2, \cdots, N. \tag{2}$$

While (2) can be solved in principle (and in practice if N is not too large), the solution in general does not fulfill the ultimate desideratum of being susceptible to representation as a simple function of the transition rates Q_{ij}. If the states are simply ordered and a transition from a given state S_i can occur only to a neighboring state S_{i-1} or S_{i+1} then the equilibrium probability π_i satisfies the following simple formula

$$\pi_i = \frac{\lambda_1\lambda_2\cdots\lambda_i}{\mu_2\mu_3\cdots\mu_{i+1}}\pi_1, \tag{3}$$

where λ_k is the transition rate from state S_k to state S_{k+1} and μ_k, the transition rate from state S_{k+1} to state S_k, and π_1 is chosen so that

$$\sum_{i=1}^{N} \pi_i = 1. \tag{4}$$

Processes of this sort, with simply ordered sets of states, are called birth and death processes. The discussion of Section 1 below deals with a class of Markov matrices Q which includes the set of birth and death matrices and allows representations analogous to (3) for the equilibrium probabilities. In Section 2 it is shown that all the matrices in this class have non-positive characteristic values and in consequence of this fact, the difference between the state probability $p_i(t)$ and the corresponding equilibrium probability π_i is majorized by a function of t and the set of initial state probabilities $p_i(0)$. In Section 3 the foregoing theory is illustrated by an anisotropic random walk.

The following notions and notations are used. Let Q be an $N \times N$ matrix. The graph $G(Q)$ associated with Q consists of vertices $V_1, V_2, \cdots, V_N$ and of

Received April 10, 1958.

directed line segments joining certain of the vertices. The directed line segment $\overline{V_jV_i}$ connecting V_j to V_i belongs to $G(Q)$ if, and only if, $Q_{ij} \neq 0$. If $k_1, k_2, \cdots, k_s$ is a sequence of indices, the product of matrix elements $Q_{k_1k_2} Q_{k_2k_3} \cdots Q_{k_{s-1}k_s}$ shall be denoted by the symbol $[k_1k_2 \cdots k_s]$. V_i shall be said to be connected to V_j if there exists a sequence of indices $k_1, k_2, \cdots, k_s$ such that $[jk_1k_2 \cdots k_si] \neq 0$ If μ is a sequence of N positive numbers, $l_2(\mu)$ shall refer to an inner product space of sequences of N complex numbers with the inner product given by

$$(x, y) = \sum_{i=1}^{N} x_i \bar{y}_i \mu_i .$$

Every $N \times N$ matrix represents an operator on $l_2(\mu)$. A matrix A is called $l_2(\mu)$-symmetric if $(Ax, y) = (x, Ay)$ for every $x \varepsilon l_2(\mu)$ and every $y \varepsilon l_2(\mu)$.

1. The equilibrium probabilities. Let μ be a sequence of N positive numbers. It is clear that a matrix A is $l_2(\mu)$-symmetric if, and only if, for every pair of indices i and j,

$$A_{ij}\mu_i = A_{ji}\mu_j . \tag{5}$$

What conditions must a matrix A satisfy in order to be $l_2(\mu)$-symmetric for some sequence μ? The answer is that the matrix is characterized by the "loop condition", i.e., the product of matrix elements around a closed loop is independent of the direction in which the loop is traversed. This statement is made precise in

THEOREM 1. *A matrix A with non-negative off-diagonal elements A_{ij} such that $A_{ij} \neq 0$ implies $A_{ji} \neq 0$ is $l_2(\mu)$-symmetric for some sequence μ of N positive numbers if, and only if, for every index i and every sequence of indices $k_1, k_2, \cdots, k_s$,*

$$[ik_1k_2 \cdots k_si] = [ik_sk_{s-1} \cdots k_1i]. \tag{6}$$

PROOF. Suppose A is $l_2(\mu)$-symmetric. A straight-forward induction based on (5) shows that (6) holds.

Now suppose that (6) holds for every sequence of indices $(k_1, k_2, \cdots k_s)$ and every index i. The proof shall be based on the construction of a μ sequence of N positive numbers such that A is $l_2(\mu)$-symmetric.

By hypothesis, if $\overline{V_iV_j}$ is in $G(A)$, then so is $\overline{V_jV_i}$. Thus $G(A)$ is the union of graphs $G(A) = G_1 + G_2 + \cdots + G_m$, which are pairwise disconnected, each one separately being, however, a connected graph.[1] Let G_i be a particular one of these subgraphs and V_i one of its vertices. Let $\mu_i > 0$ and for every $V_j \varepsilon G_i$ define

$$\mu_j = \frac{[ik_1k_2 \cdots k_sj]}{[jk_sk_{s-1} \cdots k_1i]} \mu_i , \tag{7}$$

where the chain of vertices in G_i with indices $k_1, k_2 \cdots, k_s$ connects i to j. It is clear that (6) guarantees the uniqueness of μ_j once μ_i has been chosen. For, sup-

[1] A graph G is called connected if a) in case the line $\overline{V_mV_n} \varepsilon G$ then also $\overline{V_nV_m} \varepsilon G$ and b) every pair of vertices is connected by at least one chain of lines.

pose $k_1', k_2', \cdots, k_t'$ were another chain connecting i to j. Let

$$\mu_j' = \frac{[ik_1'k_2' \cdots k_t'j]}{[jk_t'k_{t-1}' \cdots k_1'i]}\,\mu_i.$$

(6) requires that

$$[ik_1k_2 \cdots k_sj][jk_t'k_{t-1}' \cdots k_1'i] = [ik_1' \cdots k_t'j][jk_s \cdots k_1i]$$

and thus, $\mu_j = \mu_j'$. The procedure is repeated for all G_i. For the resulting sequence μ it is apparent that all $\mu_i > 0$ and that (5) is satisfied.

COROLLARY 1. *If A has non-negative off-diagonal elements and the graph $G(A)$ consists of only one connected set and* (6) *is satisfied, then there exists a unique sequence μ of positive numbers such that A is $l_2(\mu)$-symmetric and*

$$\mu_1^{-1} + \mu_2^{-1} + \cdots + \mu_N^{-1} = 1. \tag{8}$$

In addition to possessing non-negative off-diagonal elements, a Markov matrix Q has the property that

$$Q_{ii} = -\sum_{j \neq i} Q_{ji} \tag{9}$$

and thus that the sum of all of the elements in a column vanishes. A consequence of this fact is

THEOREM 2. *If a Markov matrix Q is $l_2(\mu)$-symmetric for some sequence μ satisfying* (8) *and its graph is connected, then its equilibrium probabilities π_i are given by $\pi_i = \mu_i^{-1}$.*

PROOF. By hypothesis (5) holds: $Q_{ij}\,\mu_j^{-1} = Q_{ji}\,\mu_i^{-1}$. Summation of both members with respect to i and use of (9) complete the proof. The preceding results are summarized in

THEOREM 3. *If the evolution of state probabilities of a Markov process is governed by* (1) *and transition is possible between any pair of states in one or more jumps and for every i and every set (k_α) of indices, $[i\,k_1k_2 \cdots k_s\,i] = [i\,k_sk_{s-1} \cdots k_1\,i]$, and $Q_{ij} \neq 0$ implies $Q_{ji} \neq 0$, then the equilibrium probability π_f is given by*

$$\pi_f = [fs\,s-1 \cdots 1i]\pi_i/[i12 \cdots sf], \tag{10}$$

where the initial state has been chosen arbitrarily and the set of states $S_1, S_2, \cdots, S_s$ has been picked so that the product in the denominator does not vanish. Finally, of course, (10) *together with $\sum_{k=1}^{N} \pi_k^{-1} = 1$ prescribes a unique set of equilibrium probabilities.*

2. The approach to equilibrium. If a transition rate matrix Q is $l_2(\mu)$-symmetric then it represents a symmetric operator on $l_2(\mu)$. But every characteristic value of Q is of the form $\lambda(x) = (Qx, x)/(x, x)$ for some sequence $x \in l_2(\mu)$. Therefore $\lambda(x) \leqq 0$ implies that all characteristic values of Q are non-positive. We have

$$(Qx, y) = 1/2 \sum_{i=1}^{N} \sum_{j=1}^{N} \overline{(y_i\mu_i - y_j\mu_j)}\,(Q_{ij}x_j - Q_{ji}x_i), \tag{11}$$

but $Q_{ij}\mu_i = Q_{ji}\mu_j$. Therefore,

$$(Qx, y) = -1/2 \sum_{i=1}^{N} \sum_{j=1}^{N} Q_{ji}/\mu_i \overline{(y_i\mu_i - y_j\mu_j)}\,(x_i\mu_i - x_j\mu_j),$$

and thus $\lambda(x) \leqq 0$. The result is stated in

THEOREM 4. *If a Markov matrix is $l_2(\mu)$-symmetric then all of its characteristic values are non-positive. Zero is the largest characteristic value.*

The importance of the characteristic values of Q lies in the fact that they control the approach to equilibrium. This fact is elucidated in

THEOREM 5. *Let Q be an $l_2(\mu)$-symmetric Markov matrix with connected graph. Let π be the sequence of equilibrium probabilities. Let $p(t)$ be the solution of (1) with initial value $p(0)$. If $-\lambda_M$ and $-\lambda_m$ are the largest and smallest negative characteristic values of Q then*

$$\exp(-\lambda_m t) \parallel p(0)-\pi \parallel \leqq \parallel p(t)-\pi \parallel \leqq \exp(-\lambda_M t) \parallel p(0)-\pi \parallel$$

PROOF. Letting $r(t) = p(t) - \pi$, (1) implies that

$$\frac{d \parallel r \parallel^2}{dt} = 2\left(r, \frac{dr}{dt}\right) = 2(r, Qr).$$

Since $p(t) = a\,\pi + q$, where $(q, \pi) = 0$, we have

$$a = (p(t), \pi) = \sum_{k=1}^{N} p_k(t)\ \pi_k\pi_k^{-1} = \sum_{k=1}^{N} p_k(t) = 1,$$

and thus $r(t)$ is orthogonal to π. Hence, on identifying characteristic values by their variational properties (see, for example, [2], p. 230),

$$-\lambda_m = 2 \min_{q \perp \pi} (q, Qq)/\parallel q \parallel^2 \leqq \frac{d \parallel \pi \parallel^2}{dt} \Big/ \parallel \pi \parallel^2 \leqq 2 \min_{q \perp \pi} (q, Qq)/\parallel q \parallel^2 = -\lambda_M$$

Integration and identification of $r(t)$ and the constant of integration complete the proof.

3. A homogeneous anisotropic random walk on a finite lattice. Suppose a particle moves about among the points of a finite 3-dimensional lattice and that the conditional probability that at time $t + \Delta t$ it is at point $P' = (x_1', x_2', x_3')$ having been at $P = (x_1, x_2, x_3)$ at time t is given by:

M_E	for	$x_1' = x_1 + 1$	$x_2' = x_2$	$x_3' = x_3$
M_W	"	$x_1' = x_1 - 1$	$x_2' = x_2$	$x_3' = x_3$
M_N	"	$x_1' = x_1$	$x_2' = x_2 + 1$	$x_3' = x_3$
M_S	"	$x_1' = x_1$	$x_2' = x_2 - 1$	$x_3' = x_3$
M_U	"	$x_1' = x_1$	$x_2' = x_2$	$x_3' = x_3 + 1$
M_D	"	$x_1' = x_1$	$x_2' = x_2$	$x_3' = x_3 - 1$
0		otherwise.		

Transition rates to points other than those of the lattice vanish. The resulting process is $l_2(\mu)$-symmetric and thus if π_p denotes the equilibrium probability of the particle's being at $P = (x_1, x_2, x_3)$ and π_0 that of being at the origin, then

$$\pi_p = \left(\frac{M_E}{M_W}\right)^{x_1} \left(\frac{M_N}{M_S}\right)^{x_2} \left(\frac{M_Z}{M_N}\right)^{x_3} \pi_0.$$

It is a pleasure to acknowledge the help provided through stimulating conversations with my colleagues J. Tukey and V. E. Beneš.

(*Added in proofs.*) There is considerable overlap between the work reported here and the results (particularly Theorem VIII) of the Research Announcement, "Integral Representations for Markov Transition Probabilities," by D. G. Kendall, *Bulletin A.M.S.* **64** (1958), 358–362. The notion of symmetrizability was suggested to the author, as it seems to have been to D. G. Kendall (who calls it reversability), by the spectral decomposition of Birth and Death Markov matrices effected by Kac [(1)], Ledermann and Renter [7], and Karlin and McGregor [2]. (References are to the bibliography of the above cited announcement.)

REFERENCES

[1] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, New York, 1952.

[2] F. Riesz et B. Sz. Nagy *Lecons d'Analyse Fonctionelle*, Académie des Sciences de Hongrie, Budapest, 1952.

ON SOME STATISTICAL TESTS FOR *M*TH ORDER MARKOV CHAINS[1]

By Leo A. Goodman

University of Chicago

1. Introduction and summary. Certain ψ^2 statistics were defined by Good in [8] and he stated that there is a "strong analogy" between certain functions of these statistics and some given likelihood ratio (LR) statistics appropriate for testing hypotheses concerning the order of a Markov chain. He also indicated that the analogy held when the hypothesis of "perfect randomness" is true. The present author has indicated in [12] that, for the ψ^2 statistics in [8], this analogy (i.e., the asymptotic equivalence of the corresponding statistics) does not hold under some more general conditions when the "perfect randomness" hypothesis is not true. It will be seen herein that certain functions of a modified form of the ψ^2 statistics are asymptotically equivalent to certain LR statistics in the more general case when the hypothesis $H(P_m)$ that the positively regular Markov chain (see [2]) is governed by a completely specified system P_m of mth order transition probabilities is true. Also, certain functions of a different modified form of the ψ^2 statistics will be seen to be asymptotically equivalent to certain LR statistics in the case when the hypothesis H_m that the positively regular Markov chain is of order m is true. These results are helpful in determining the asymptotic distributions of various statistics and the null hypotheses that can be tested with a given statistic. For example, if a given statistic G is asymptotically equivalent, under $H(P_1)$, to the LR statistic L for testing the null hypothesis $H(P_1)$ within the alternate hypothesis H_2, then the asymptotic distribution, under $H(P_1)$, of G will be χ^2 with a known number of degrees of freedom (i.e., with a known expectation); G can be used directly to test $H(P_1)$ within H_2 (if G is sensitive to these hypotheses), although the asymptotic distribution, under H_2, of G may differ, in a certain sense, from that of L (see Section 6 in [1]). However, if a given statistic ΔG is asymptotically equivalent, under $H(P_1)$, to the LR statistic ΔL for testing the null hypothesis H_1 within the alternate hypothesis H_2 (i.e., if there is an "ostensible analogy" between ΔG and ΔL), but this asymptotic equivalence does not hold under some more general conditions (e.g., under H_1), then ΔG can not be used to test H_1 within H_2; the asymptotic distribution, under $H(P_1)$, of ΔG will be χ^2, but the asymptotic disbution, under the null hypothesis H_1, will not be χ^2, and furthermore the expectation, under H_1, of G can approach infinity (see [15]).

The present author has indicated in [15] that certain functions of a modified

Received June 3, 1958; revised September 24, 1958.

[1] Research carried out at the Statistical Research Center, University of Chicago, under the sponsorship of the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

I am indebted to I. J. Good for bringing [18] to my attention, and to S. Ghurye for helping with the Russian translation.

form of the ψ^2 statistics, which were investigated by Stepanow in [18] and which are computed for a specified P_1, are asymptotically equivalent, under $H(P_1)$, to certain LR statistics, but that they will not be equivalent under H_1. Although it is stated in [18] that the results presented there can be applied to the solution of the problem of testing the null hypothesis H_1, it is shown in [15] that none of the statistics in [18] can be used directly to test this composite hypothesis. In the present paper, it will be seen that a statistic based on a different modified form of ψ^2, as well as certain other statistics described herein, will be asymptotically equivalent, under H_1, to a certain LR statistic and can be used to test the null hypothesis H_1.

The ψ^2 statistics defined by Good in [10] are more general than those given in [3], [8], [18]. Besides studying the relation between these statistics and the LR statistics, we shall also discuss certain conjectures proposed in [10] concerning the asymptotic distributions of these statistics, which were investigated by Billingsley [4] for the cases H_0 and H_1 (the author mentions that a more general result for H_m $(m \geqq 0)$ can be obtained using similar methods) and independently, using different methods, by the present author [14] for the case H_m $(m \geqq 0)$ when the transition probabilities are all positive (this author also mentions that a more general result can be obtained by similar methods). The ψ^2 and LR statistics defined in [10] and [8], as well as some related statistics developed in the present paper, will be generalized further herein, and the asymptotic distributions of these generalized statistics will be investigated. This investigation leads to generalizations of the asymptotic distributions obtained by Good [8], Billingsley [3] [4], and the present author [13] [14], and it helps to clarify the relation between the various statistics.

The different asymptotically equivalent forms of various statistics presented here make it possible for the statistician to choose whichever form he finds preferable both from the computational point of view and also from some other viewpoints (see [1], [5], [11]).

2. The first order chain. Let $\{X_1, X_2, \cdots, X_n\}$ be an observed sequence from a stochastic process. It will be convenient to deal herein with a circularized sequence of observations obtained by regarding the first observation X_1 as immediately following the nth observations X_n (see [8], [12]). In this case, the frequency $f(y_s)$ of the s consecutive observations (i.e., the s-tuple) $y_s = (u_1, u_2, \cdots, u_s)$ in the circularized sequence will be such that $\sum_{u_1} f(y_s) = \sum_{w_s} f(w_s) = f(w_{s-1})$, where $w_s = (w_1, w_2, \cdots, w_s)$, $(w_1, w_2, \cdots, w_{s-1}) = (u_2, u_3, \cdots, u_s) = w_{s-1}$, and $f(w_{s-1})$ is the frequency of the $(s-1)$-tuple w_{s-1} in the circularized sequence. A method of modifying results obtained for circularized sequences so that they can be applied to noncircularized sequences has been given in [12]; results for circularized sequences can not in general be applied directly to noncircularized sequences (see [12] and Corrigenda to [8]).

The following result has been presented in [18]: Consider an observed sequence $\{X_1, X_2, \cdots, X_n\}$ from a positively regular Markov chain with constant transition probability matrix $P_1 = (p_{ij})$, where the possible states are $1, 2, \cdots, a$.

Let p_i' denote the stationary probabilities, and let k_s be the number of s-tuples that are possible given P_1 ; e.g., if all $p_{ij} > 0$, then $k_s = a^s$. Let $\psi^2_{1,s} = \sum_{y_s} [f(y_s) - f_1(y_s)]^2/f_1(y_s)$, where $f_1(y_s) = np'_{u_1}\prod_{i=1}^{s-1} p_{u_i u_{i+1}}$ is the expected value (asymptotically) of $f(y_s)$ in the new sequence of length n given $H(P_1)$, and where the summation is taken over the k_s values of y_s where $f_1(y_s) > 0$. (The $f_1(y_s)$ given above is not the exact expected value, but is an asymptotic approximation; similar asymptotic approximations for expected values will be used throughout). Then (a) the statistics $\Delta\psi^2_{1,s} = \psi^2_{1,s} - \psi^2_{1,s-1}$ $(s \geqq 2)$ are asymptotically distributed $(n \to \infty)$ as χ^2 with $\Delta k_s = k_s - k_{s-1}$ degrees of freedom (d.f.), and (b) the $\Delta^2\psi^2_{1,s} = \psi^2_{1,s} - 2\psi^2_{1,s-1} + \psi^2_{1,s-2}$ for $s \geqq 3$ are asymptotically independent and distributed as χ^2 with $\Delta^2 k_s = k_s - 2k_{s-1} + k_{s-2}$ d.f.

We shall now introduce ϕ^2 statistics, which are related to, but different from, the ψ^2 statistics. Let $\phi^2_{1,s} = 2\sum_{y_s} f(y_s) \log [f(y_s)/f_1(y_s)]$. Then, for $s \geqq 2$,

$$\begin{aligned}\Delta\phi^2_{1,s} &= \phi^2_{1,s} - \phi^2_{1,s-1} \\ &= 2\sum_{y_s} f(y_s) \log [f(y_s)/f_1'(y_s)] \\ &= L_{1,s},\end{aligned}$$

where $y_{s-1} = (u_1, u_2, \cdots, u_{s-1})$, and $f_1'(y_s) = f(y_{s-1})p_{u_{s-1}u_s}$ is the expected value (asymptotically) of $f(y_s)$ in a new sequence of length n given $f(y_{s-1})$ and $H(P_1)$. For $s \geqq 3$,

$$\begin{aligned}\Delta^2\phi^2_{1,s} &= \Delta\phi^2_{1,s} - \Delta\phi^2_{1,s-1} = \Delta L_{1,s} \\ &= 2\sum_{y_s} f(y_s) \log [f(y_s)/\bar{f}_{s-2}(y_s)] \\ &= M_{s-2,s}\end{aligned}$$

where $w_{s-2} = (w_1, w_2, \cdots, w_{s-2})$, $\bar{P}_{s-2}$ is the maximum likelihood estimate of the $(s-2)$th order transition probability matrix when H_{s-2} is true (see [14]), and $\bar{f}_{s-2}(y_s) = f(y_{s-1})f(w_{s-1})/f(w_{s-2})$ is the expected value (asymptotically) of $f(y_s)$ in a new sequence of length n given $f(y_{s-1})$ and $H(\bar{P}_{s-2})$. Let

$$K_s = 2\sum_{y_s} f(y_s) \log f(y_s),$$

$$K_{1,2} = 2\sum_{y_2} f(y_2) \log p(y_2),$$

and

$$K_{1,1} = 2\sum_{y_1} f(y_1) \log np'(y_1),$$

where $p(y_2) = p_{u_1u_2}$ and $p'(y_1) = p'_{u_1}$. Then $\phi^2_{1,s} = K_s - (s-1)K_{1,2} - K_{1,1}$, $\Delta\phi^2_{1,s} = \Delta K_s - K_{1,2}$, and $\Delta^2\phi^2_{1,s} = \Delta^2 K_s$. It can be seen that, given $H(P_1)$, the statistics $\phi^2_{1,s}$ are asymptotically equivalent to the $\psi^2_{1,s}$, and that (a) the $\Delta\phi^2_{1,s} = \Delta K_s - K_{1,2} = L_{1,s}$ $(s \geqq 2)$ are asymptotically χ^2 with Δk_s d.f., and (b) the $\Delta^2\phi^2_{1,s} = \Delta L_{1,s} = \Delta^2 K_s = M_{s-2,s}$ are asymptotically independent and distributed as χ^2 with $\Delta^2 k_s$ d.f. (e.g., see methods in [1], [4], [14]).

Let

$$G_{1,s} = \sum_{\underline{y}_s} [f(\underline{y}_s) - f_1'(\underline{y}_s)]^2 / f_1'(\underline{y}_s),$$

and

$$F_{s-2,s} = \sum_{\underline{y}_s} [f(\underline{y}_s) - f_{s-2}(\underline{y}_s)]^2 / f_{s-2}(\underline{y}_s).$$

Then it can be seen (with methods in [1], [4]) that, given $H(P_1)$, the $G_{1,s}$ are asymptotically equivalent to $L_{1,s} = \Delta\phi^2_{1,s}$ and thus to $\Delta\psi^2_{1,s}$. Hence, the $\Delta G_{1,s} = G_{1,s} - G_{1,s-1}$ are asymptotically equivalent, given $H(P_1)$, to $\Delta^2\psi^2_{1,s}$. Also, the $F_{s-2,s}$ are asymptotically equivalent, given $H(P_1)$, to $M_{s-2,s} = \Delta^2\phi^2_{1,s}$ and thus to $\Delta^2\psi^2_{1,s}$.

The statistics $L_{1,s}$ are the LR statistics for testing the null hypothesis $H(P_1)$ within the alternate hypothesis H_{s-1} (see [2]). Although $G_{1,s}$ and $\Delta\psi^2_{1,s}$ are asymptotically equivalent, given $H(P_1)$, to $L_{1,s}$, it can be seen that, in the case where $H(P_1)$ is not true, $G_{1,s}$ and $\Delta\psi^2_{1,s}$ are asymptotically equivalent to $L_{1,s}$ only in the special sense that the usual χ^2 goodness of fit statistic for the standard test of a simple null hypothesis concerning a multinomial distribution is considered to be asymptotically equivalent (even when this null hypothesis is not true) to the usual LR statistic for this hypothesis (see Section 6 of [1]); i.e., if the null hypothesis $H(P_1)$ is not true, these statistics will be asymptotically equivalent only if the true hypothesis approaches, in a certain sense, $H(P_1)$ at a sufficiently fast rate as $n \to \infty$. The relative advantages and disadvantages of $L_{1,s}$, $\Delta\psi^2_{1,s}$, and $G_{1,s}$ as tests of $H(P_1)$ will not be discussed here, since such discussions for somewhat related problems appear in [1], [5], and [11].

The statistics $M_{s-2,s}$ are the LR statistics for testing the null hypothesis H_{s-2} within H_{s-1}, and their asymptotic distribution, under H_{s-2}, is χ^2 with $\Delta^2 k_s$ d.f. (see [8], [16]). Although $\Delta^2\psi^2_{1,s}$ and $\Delta G_{1,s}$ are asymptotically equivalent, given $H(P_1)$, to $M_{s-2,s}$, it can be seen that the former statistics are not asymptotically equivalent, given H_{s-2}, to $M_{s-2,s}$, and their asymptotic distribution will depend on P_1 (which is used in the computation of $\Delta^2\psi^2_{1,s}$ and $\Delta G_{1,s}$) and on the particular system P'_{s-2} of transition probabilities that is true when H_{s-2} is true (P'_{s-2} can be viewed as a particular $a^{s-2} \times a^{s-2}$ matrix of transition probabilities describing a given Markov chain of order $s-2$); e.g., for some P_1 that differ from P'_{s-2}, the statistics $\Delta^2\psi^2_{1,s}$ will converge in probability, given $H(P'_{s-2})$, to infinity even though the null hypothesis H_{s-2} is true (see [15]). Thus, the asymptotic distribution, given H_{s-2}, of the statistics $\Delta^2\psi^2_{1,s}$ and $\Delta G_{1,s}$ will depend on unknown values of the parameters; these statistics can not be used to test the null hypothesis H_{s-2} in the same simple manner as when the LR statistic $M_{s-2,s}$ is used. However, it can be seen that $F_{s-2,s}$ is equivalent, given H_{s-2}, to $M_{s-2,s}$; and thus can be used to test H_{s-2} within H_{s-1} (see [12], [13]).

3. The general case. Stepanow [18] mentions that, given $H(P_1)$, the asymptotic independence of the $\Delta^2\psi^2_{1,s}$ statistics leads to the fact that the statistics $\Delta\psi^2_{1,s}$ —

$\Delta\psi^2_{1,s}$ $(s > t \geqq 2)$ are asymptotically χ^2 with $\Delta k_s - \Delta k_t$ d.f. It can also be seen, given H_{s-2}, that the $\Delta^2 K_j$ statistics (for $j \geqq s$) are asymptotically independent (see [4], [14]); thus the $\Delta K_s - \Delta K_t$ statistics, given H_{t-1}, are asymptotically χ^2 with $\Delta k_s - \Delta k_t$ d.f. (see [8]). To test H_{t-1} within H_{s-1}, the $\Delta K_s - \Delta K_t$ are the LR statistics (see [8]). The $\Delta\psi^2_{1,s} - \Delta\psi^2_{1,t}$, given H_{t-1}, are not asymptotically equivalent to the LR statistics, and can not serve as a test of H_{t-1} (see [15]).

We have that $\Delta K_s - \Delta K_t = 2\sum_{y_s} f(y_s) \log (f(y_s)/\hat{f}_{t-1}(y_s)) = M_{t-1,s}$, where $y_t = (w_1, w_2, \cdots, w_t) = (u_{s-t+1}, u_{s-t+2}, \cdots, u_s)$, $y_{t-1} = (w_1, w_2, \cdots, w_{t-1})$, and $\hat{f}_{t-1}(y_s) = f(y_{s-1})f(y_t)/f(y_{t-1})$ is the expected value (asymptotically) of $f(y_s)$ in a new sequence of length n given $f(y_{s-1})$ and $H(\bar{P}_{t-1})$. Let $F_{t-1,s} = \sum_{y_s} [f(y_s) - \hat{f}_{t-1}(y_s)]^2/\hat{f}_{t-1}(y_s)$. Then, it can be seen that, given H_{t-1}, the statistics $F_{t-1,s}$ are asymptotically equivalent to $M_{t-1,s} = \Delta K_s - \Delta K_t = \Delta\phi^2_{1,s} - \Delta\phi^2_{1,t}$.

Let

$$\hat{\psi}^2_{t-1,s} = \sum_{y_s} [f(y_s) - \hat{f}_{t-1}(y_s)]^2/\hat{f}_{t-1}(y_s),$$

where (for $s > t \geqq 1$)

$$y_t = (u_1, u_2, \cdots, u_t),\ y_{t,i} = (u_i, u_{i+1}, \cdots, u_{i+t-1}),$$

$$y_{t-1,i} = (u_i, u_{i+1}, \cdots, u_{i+t-2}),$$

and

$$\hat{f}_{t-1}(y_s) = f(y_{t-1}) \prod_{i=1}^{s-t+1} f(y_{t,i})/f(y_{t-1,i})$$

is the expected value (asymptotically) of $f(y_s)$ in a new sequence of length n given $f(y_{t-1})$ and $H(\bar{P}_{t-1})$. Then, given H_{t-1}, the statistics $\hat{\psi}^2_{t-1,s}$ are asymptotically equivalent to

$$\hat{\phi}^2_{t-1,s} = 2\sum_{y_s} f(y_s) \log [f(y_s)/\hat{f}_{t-1}(y_s)] = K_s - K_{t-1} - (s - t + 1)(\Delta K_t).$$

Since $\Delta\hat{\phi}^2_{t-1,s} = \Delta K_s - \Delta K_t$, the $\Delta\hat{\phi}^2_{t-1,s}$ or the $\Delta\hat{\psi}^2_{t-1,s}$, as well as the $F_{t-1,s}$, can be used to test H_{t-1} within H_{s-1}.

Consider now the hypothesis $H(P_t)$ that the positively regular Markov chain is of the tth order and governed by the system of transition probabilities P_t (see [2]). Let $\Pr\{u_{t+1} \mid u_1, u_2, \cdots, u_t\} = p_{u_1 u_2 \cdots u_{t+1}} = p(y_{t+1})$ denote the transition probability that the j-th observation in the sequence will be u_{t+1}, given that the $(j-1)$-th, $(j-2)$-th, $\cdots$, $(j-t)$-th observations were $u_t, u_{t-1}, u_{t-2}, \cdots, u_1$ respectively, and let $p'(y_t)$ denote the stationary (absolute) probability for y_t. Let $\phi^2_{t,s} = 2\sum_{y_s} f(y_s) \log[f(y_s)/f_t(y_s)]$, where (for $s > t \geqq 1$) $f_t(y_s) = np'(y_t) \prod_{i=1}^{s-t} p_{u_i u_{i+1} \cdots u_{i+t}}$ is the expected value (asymptotically) of $f(y_s)$ in a new sequence of length n given $H(P_t)$. Then

$$\Delta\phi^2_{t,s} = \phi^2_{t,s} - \phi^2_{t,s-1} = 2\sum_{y_s} f(y_s) \log (f(y_s)/f'_t(y_s)) = L_{t,s},$$

where $f_t'(y_s) = f(y_{s-1})p_{u_{s-t}u_{s-t+1}\cdots u_s}$ is the expected value (asymptotically) of $f(y_s)$ in a new sequence of length n given $f(y_{s-1})$ and $H(P_t)$; and

$$\Delta^2\phi_{t,s}^2 = \Delta\phi_{t,s}^2 - \Delta\phi_{t,s-1}^2 = \Delta L_{t,s}$$
$$= 2\sum_{y_s} f(y_s)\log[f(y_s)/f_{s-2}(y_s)] = M_{s-2,s} = \Delta^2\phi_{1,s}^2 = \Delta^2\phi_s^2.$$

Since

$$\phi_{t,s}^2 = K_s - (s-t)K_{t,t+1} - K_{t,t},$$

where

$$K_{t,t+1} = 2\sum_{y_{t+1}} f(y_{t+1})\log p(y_{t+1})$$
$$K_{t,t} = 2\sum_{y_t} f(y_t)\log np'(y_t),$$

we see that

$$\Delta\phi_{t,s}^2 = \Delta K_s - K_{t,t+1}.$$

Let $\psi_{t,s}^2 = \sum_{y_s}[f(y_s) - f_t(y_s)]^2/f_t(y_s)$, and $G_{t,s} = \sum_{y_s}[f(y_s) - f_t'(y_s)]^2/f_t'(y_s)$. Then it can be seen that, given $H(P_t)$, the $\Delta\psi_{t,s}^2$, $\Delta\phi_{t,s}^2 = L_{t,s}$ and $G_{t,s}$ are all asymptotically equivalent, and each is asymptotically distributed as χ^2 with Δk_s d.f.

The reader will note that k_s and the statistics mentioned in the preceding paragraph depend on P_t. In the case where the null hypothesis tested is H_t, the particular statistics mentioned earlier herein appropriate for such a test do not depend on P_t, but their distribution does since k_s does. If the null hypothesis to be tested is H_t (and P_t is not specified), then the value of k_s to be used can be estimated consistently from the observed number of y_s where $f(y_s) > 0$. (In the general case where the chain may contain some transient states, but where only the recurrent states are of interest and the transient states are not, the condition $f(y_s) > 0$ should be applied to the sequence obtained by omitting all observations before the first one that is in a recurrent state (see [4]); the recurrent states can also be estimated consistently from the observed sequence.)

4. The distribution of the $\hat{\psi}^2$ and ψ^{+2} statistics. Conjectures concerning the asymptotic distribution of $\hat{\psi}_{t,s}^2$ and $\psi_{t,s}^2$ were proposed in [10] and modified forms of these conjectures were proved in [4] and [14]. We now present the following generalization of these results:

(A) The asymptotic distribution of $\hat{\psi}_{t,s}^2$, given H_t, is $*_{\lambda=1}^{s-t-1} K_{g(\lambda)}(x/\lambda)$, where $*$ denotes convolution, $g(\lambda) = \Delta^2 k_{(s+1-\lambda)}$, and $K_g(x)$ is the χ^2 distribution with g degrees of freedom.

(B) Let $\psi_{t,s}^{+2} = \sum_{y_s}[f(y_s) - f_t^+(y_s)]^2/f_t^+(y_s)$, where

$$f_t^+(y_s) = f(y_t)\prod_{i=1}^{s-t} p_{u_iu_{i+1}\cdots u_{i+t}}$$

is the expected value (asymptotically) of $f(y_s)$ in a new sequence of length n

given $H(P_t)$ and $f(\underline{y}_t)$. Then the asymptotic distribution ($n \to \infty$) of $\psi^{*2}_{t,s}$, given $H(P_t)$, is $*_{\lambda=1}^{s-t-1} K_{\varphi(\lambda)}(x/\lambda) * K_{h(t)}[x/(s-t)]$, where $h(t) = \Delta k_{t+1}$.

Statement (A) can be seen to follow from the fact that $\hat{\psi}^2_{t,s}$ is asymptotically equivalent, given H_t, to

$$\phi^2_{t,s} = K_s - K_t - (s-t)\Delta K_{t+1} = \sum_{j=1}^{s-t-1} j\Delta^2 K_{(s+1-j)}, \tag{4.1}$$

where the statistics $\Delta^2 K_s$ (for $s > t+1$) are asymptotically independent (see [4], [14]). Statement (B) can be seen to follow from the fact that $\psi^{*2}_{t,s}$ is asymptotically equivalent, under $H(P_t)$, to

$$\begin{aligned} \phi^{*2}_{t,s} &= 2\sum_{\underline{y}_s} f(\underline{y}_s) \log [f(\underline{y}_s)/f^*_t(\underline{y}_s)] \\ &= K_s - K_t - (s-t)K_{t,t+1} = \phi^2_{t,s} + (s-t)\Delta\phi^2_{t,t+1} \end{aligned} \tag{4.2}$$

where $\phi^2_{t,s}$ and $\Delta\phi^2_{t,t+1}$ are asymptotically independent (see [4], [14]). We also note that the asymptotic distribution of $\psi^{*2}_{t,s}$ (or $\phi^{*2}_{t,s}$) is different from that of $\psi^2_{t,s}$; but the asymptotic distributions, given $H(P_t)$, of $\Delta\psi^2_{t,s}$ and $\Delta\psi^{*2}_{t,s}$ are identical since $\Delta\phi^2_{t,s} = \Delta\phi^{*2}_{t,s}$.

The results presented here were for $t \geq 1$. The case where $t = 0$ can be treated in a similar fashion (see [3], [9], [10], [14]).

5. Some generalized statistics and their distributions. From (4.1) we see that, for $s = t+2$, $\phi^2_{t,s}$ and $\hat{\psi}^2_{t,s}$ are asymptotically equivalent, given H_t, to $\Delta^2 K_s = M_{t,s}$, the LR statistic for testing the null hypothesis H_t within the alternate hypothesis H_{s-1} (i.e., $M_{t,s} = -2 \log \lambda_{t,s-1}$, where $\lambda_{t,s-1}$ is the ratio of the maximum likelihood given H_t to that given H_{s-1}). This relationship between $M_{t,s}$ and $\hat{\psi}^2_{t,s}$ (and $\phi^2_{t,s}$) does not hold for $s > t+2$. Also, for $s = t+1$, $\phi^{*2}_{t,s}$ and $\psi^{*2}_{t,s}$ are asymptotically equivalent, under $H(P_t)$, to $\Delta\phi^2_{t,s} = L_{t,s}$, the LR statistic for testing the null hypothesis $H(P_t)$ within H_{s-1}. This relationship between $L_{t,s}$ and $\psi^{*2}_{t,s}$ (and $\phi^{*2}_{t,s}$) does not hold for $s > t+1$. We shall now present, for $s \geqq t+2$, a generalized statistic $\phi^2_{t,s;r}$ that will include both $\phi^2_{t,s}$ and $M_{t,s}$ as special cases, and a statistic $\hat{\psi}^2_{t,s;r}$ that will be asymptotically equivalent, given H_t, to $\phi^2_{t,s;r}$. Also, we shall present, for $s \geqq t+1$, a generalized statistic $\phi^{*2}_{t,s;r}$ that will include both $\phi^{*2}_{t,s}$ and $L_{t,s}$ as special cases, and a statistic $\psi^{*2}_{t,s;r}$ that will be asymptotically equivalent, given $H(P_t)$, to $\phi^{*2}_{t,s;r}$. Finally, a generalized statistic $M_{t,s;r}$ (different from $\phi^2_{t,s;r}$) will be presented that will include $M_{t,s}$ as a special case, and the asymptotic distribution of each of these generalized statistics will be investigated.

Let

$$\begin{aligned} \phi^2_{t,s;r} &= 2\sum_{\underline{y}_s} f(\underline{y}_s) \log [f(\underline{y}_s)/\hat{f}_{t;r}(\underline{y}_s)], \\ \hat{\psi}^2_{t,s;r} &= \sum_{\underline{y}_s} [f(\underline{y}_s) - \hat{f}_{t;r}(\underline{y}_s)]^2/\hat{f}_{t;r}(\underline{y}_s), \\ \phi^{*2}_{t,s;r} &= 2\sum_{\underline{y}_s} f(\underline{y}_s) \log [f(\underline{y}_s)/f^*_{t;r}(\underline{y}_s)], \end{aligned}$$

and

$$\psi^{*2}_{t,s;r} = \sum_{y_s} [f(y_s) - f^*_{t;r}(y_s)]^2 / f^*_{t;r}(y_s),$$

where

$$\hat{f}_{t;r}(y_s) = f(y_{t+r}) \prod_{i=1}^{s-t-r} f(y_{t+1,i+r}) / f(y_{t,i+r}),$$

$$f^*_{t;r}(y_s) = f(y_{t+r}) \prod_{i=1}^{s-t-r} p_{u_{i+r}u_{i+r+1}\cdots u_{i+r+t}},$$

$$y_{t+r} = (u_1, u_2, \cdots, u_{t+r}),\ y_{t+1,i+r} = (u_{i+r}, u_{i+r+1}, \cdots, u_{i+r+t}),$$

and

$$y_{t,i+r} = (u_{i+r}, u_{i+r+1}, \cdots, u_{i+r+t-1}) \quad \text{for } 0 \leqq r < s - t.$$

Then for $r = 0$,

$$\hat{\phi}^2_{t,s;r} = \hat{\phi}^2_{t,s}, \quad \hat{\psi}^2_{t,s;r} = \hat{\psi}^2_{t,s}, \quad \phi^{*2}_{t,s;r} = \phi^{*2}_{t,s}, \quad \psi^{*2}_{t,s;r} = \psi^{*2}_{t,s};$$

for $r = s - t - 1$,

$$\hat{\phi}^2_{t,s;r} = M_{t,s}, \quad \hat{\psi}^2_{t,s;r} = F_{t,s}, \quad \phi^{*2}_{t,s;r} = L_{t,s}, \quad \psi^{*2}_{t,s;r} = G_{t,s}.$$

Let

$$M_{t,s;r} = 2 \sum_{y_s} f(y_s) \log [f(y_s)/\hat{f}_{t;r}(y_s)]$$

and

$$F_{t,s;r} = \sum_{y_s} [f(y_s) - \hat{f}_{t;r}(y_s)]^2 / \hat{f}_{t;r}(y_s),$$

where

$$\hat{f}_{t;r}(y_s) = f(y_{s-1-r}) f(w_{t+1+r}) / f(w_t),$$

$$w_{t+1+r} = (w_1, w_2, \cdots, w_{t+1+r}) = (u_{s-t-r}, u_{s-t-r+r}, \cdots, u_s),$$

and

$$w_t = (w_1, w_2, \cdots, w_t) \qquad \text{for } 0 \leqq r \leqq s - t - 2.$$

Then for $r = 0$, $M_{t,s;r} = M_{t,s}$ and $F_{t,s;r} = F_{t,s}$; for $r = s - t - 2$, $M_{t,s;r} = M_{t,s}$ and $F_{t,s;r} = \bar{F}_{t,s}$, where $\bar{F}_{t,s}$ is $F_{t,s}$ computed for the sequence $\{X_n, X_{n-1}, \cdots, X_1\}$ circularized rather than for the sequence $\{X_1, X_2, \cdots, X_n\}$. The statistic $M_{t,s;r}$ (or $F_{t,s;r}$) can be seen to be the sum of the LR statistics (or the goodness of fit statistics) for testing "independence" in each of k_t "contingency tables" (i.e., $M_{t,s;r}$ is the product of the ratios of the maximum likelihood given independence to that given "nonindependence" in each "contingency table" when normed ln the usual way; viz., -2 times the log of this product) obtained by "splitting" each s-tuple y_s into a $(s - t - 1 - r)$-tuple $(u_1, u_2, \cdots, u_{s-t-1-r})$,

a t-tuple $(u_{s-t-r}, u_{s-t-r+1}, \cdots, u_{s-r-1})$, and a $(1 + r)$-tuple $(u_{s-r}, u_{s-r+1}, \cdots, u_s)$; for each t-tuple a "contingency table" can be formed where the "expected value" of the observed cell entry $f(u_s)$ is $\hat{f}_{t;r}(u_s)$ under the assumption of independence in the table (see [13], [14]).

We shall now present simple derivations for the asymptotic distributions of the generalized statistics using a similar approach to that described in Section 4. We have that

$$\hat{\phi}^2_{t,s;r} = K_s - K_{t+r} - (s - t - r)\Delta K_{t+1} = \sum_{j=1}^{s-t-1} d(j)\Delta^2 K_{(s+1-j)}$$

where

$$d(j) = \begin{cases} j & \text{for } 0 \leqq j \leqq s - t - r \\ (s - t - r) & \text{for } s - t - r \leqq j \leqq s - t - 1. \end{cases}$$

Therefore, the asymptotic distribution $(n \to \infty)$ of $\hat{\phi}^2_{t,s;r}$, given H_t, is

$$\underset{\lambda=1}{\overset{s-t-1}{*}} K_{g(\lambda)}[x/d(\lambda)] = \underset{\lambda=1}{\overset{s-t-r-1}{*}} K_{g(\lambda)}(x/\lambda) * K_{h(t+r)-h(t)}[x/(s - t - r)].$$

This will also be the asymptotic distribution of $\hat{\psi}^2_{t,s;r}$, under H_t, since $\hat{\psi}^2_{t,s;r}$ and $\hat{\phi}^2_{t,s;r}$ are asymptotically equivalent under H_t.

By a similar approach, we see that

$$M_{t,s;r} = K_s - K_{s-1-r} - (K_{t+1+r} - K_t) = \sum_{j=1}^{s-t-1} c(j)\Delta^2 K_{(s+1-j)},$$

where

$$c(j) = \begin{cases} j & \text{for } 0 \leqq j \leqq v \\ v & \text{for } v \leqq j \leqq s - t - v \\ (s - t - j) & \text{for } s - t - v \leqq j \leqq s - t - 1, \end{cases}$$

and $v = \min[r + 1, s - t - r - 1]$. Therefore, the asymptotic distribution $(n \to \infty)$ of $M_{t,s;r}$ (and $F_{t,s;r}$), given H_t, is

$$\underset{\lambda=1}{\overset{s-t-1}{*}} K_{g(\lambda)}[x/c(\lambda)]$$

(see [14$_1$]).

We also see that

$$\phi^{*2}_{t,s;r} = K_s - K_{t+r} - (s - t - r)K_{t,t+1} = \hat{\phi}^2_{t,s;r} + (s - t - r)\Delta\phi^2_{t,t+1}.$$

Therefore the asymptotic distribution of $\phi^{*2}_{t,s;r}$ (and $\psi^{*2}_{t,s;r}$), given $H(P_t)$, is

$$\underset{\lambda=1}{\overset{s-t-1}{*}} K_{g(\lambda)}[x/d(\lambda)] * K_{h(t)}[x/(s - t - r)]$$

$$= \underset{\lambda=1}{\overset{s-t-r-1}{*}} K_{g(\lambda)}(x/\lambda) * K_{h(t+r)}[x/(s - t - r)].$$

If P_t is not completely specified but is a function $P_t(\alpha)$ of a vector parameter $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_v)$ that ranges over an open subset of v-dimensional Euclidean space, and if this function satisfies certain regularity conditions (see [4]), then it can be seen that the statistic $\phi^{*2}_{t,s;r}(\hat{\alpha})$ (or $\psi^{*2}_{t,s;r}(\hat{\alpha})$), obtained by replacing P_t by its maximum likelihood estimate $P_t(\hat{\alpha})$ in the computation of $\phi^{*2}_{t,s;r}$ (or $\psi^{*2}_{t,s;r}$), will have the same asymptotic distribution as $\phi^{*2}_{t,s;r}$ except that the degrees of freedom $h(t+r)$ should be replaced by $h(t+r) - v$. If P_t is the transition probability matrix for a completely unspecified positively regular tth order Markov chain, then $v = \Delta k_{t+1} = h(t)$, $\phi^{*2}_{t,s;r}(\hat{\alpha}) = \hat{\phi}^2_{t,s;r}$, $\psi^{*2}_{t,s;r}(\hat{\alpha}) = \hat{\psi}^2_{t,s;r}$, and the asymptotic distribution, under H_t, of these statistics was given earlier herein. This result is closely related to and generalizes the asymptotic distributions in [4] for $r = 0$ and $s - t - 1$ when $t = 0$ or 1.

This investigation of the asymptotic distributions of various generalized statistics, under particular null hypotheses, indicates that each of these statistics is asymptotically equivalent (under a particular hypothesis) to a weighted sum of the LR statistics $\Delta^2 K_s$, $\Delta^2 K_{s-1}$, $\cdots$, $\Delta^2 K_{t+2}$, $\Delta\phi^2_{t,t+1}$ (or $\Delta\phi^2_{t,t+1}(\hat{\alpha})$, where $\phi^2_{t,t+1}(\hat{\alpha})$ is defined as $\phi^2_{t,t+1}$ with P_t replaced by $P_t(\hat{\alpha})$). The particular generalized statistic that will be appropriate for a given problem will depend in part on the appropriate weighting of the LR statistics, which will in turn depend on the specific null and alternate hypotheses considered (see [4], [14]).

In closing, we point out that the asymptotic mean values of $\hat{\phi}^2_{t,s;r}$ and $M_{t,s;r}$ under H_t, and of $\phi^{*2}_{t,s;r}$ under $H(P_t)$, can be computed directly by reference to the decomposition of the various statistics in terms of the K's. When all transition probabilities are positive, these mean values are $a^{t+r}(a^{s-t-r} - 1) - (s - t - r)a^t(a - 1)$, $a^t(a^{s-t-1-r} - 1)(a^{1+r} - 1)$, and $a^{t+r}(a^{s-t-r} - 1)$ for $\hat{\phi}^2_{t,s;r}$, $M_{t,s;r}$ and $\phi^{*2}_{t,s;r}$, respectively, and they can be given some interpretation in terms of the asymptotic mean values of certain corresponding LR statistics computed from a set of a^t (and a^{t+r}) independent "contingency tables" (see [13], [14]).

REFERENCES

[1] T. W. Anderson and Leo A. Goodman, "Statistical inference about Markov chains," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 89–110.

[2] M. S. Bartlett, "The frequency goodness of fit test for probability chains," *Proc. Cambridge Philos. Soc.*, Vol. 47 (1951), pp. 86–95.

[3] P. Billingsley, "Asymptotic distributions of two goodness of fit criteria," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1123–9.

[4] P. Billingsley, "On testing Markov chains," presented at the meetings of the Institute of Mathematical Statistics, Sept. 10–13, 1957.

[5] W. G. Cochran, "The χ^2-test of goodness of fit," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 315–45.

[6] R. Dawson and I. J. Good, "Exact Markov probabilities from oriented linear graphs," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 946–56.

[7] C. Derman, "Some asymptotic distribution theory for Markov chains with a denumerable number of states," *Biometrika*, Vol. 43 (1956), pp. 285–94.

[8] I. J. Good, "The likelihood ratio test for Markov chains," *Biometrika*, Vol. 42 (1955), pp. 531–3; "Corrigenda," Vol. 44 (1957), p. 301.

[9] I. J. Good, "On the weighted combination of significance tests," *J. Roy. Stat. Soc. Series B (Methodological)*, Vol. 17 (1955), pp. 264–5.

[10] I. J. Good, "Review of P. Billingsley's 'Asymptotic distributions of two goodness of fit criteria'," *Math. Reviews*, Vol. 18 (1957), p. 607.

[11] I. J. Good, "Saddle point methods for the multinomial distribution," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 861–81.

[12] Leo A. Goodman, "Simplified runs tests and likelihood ratio tests for Markov chains," *Biometrika*, Vol. 45 (1958), pp. 181–97.

[13] Leo A. Goodman, "Exact probabilities and asymptotic relationships for some statistics from *m*th order Markov chains," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 476–90.

[14] Leo A. Goodman, "Asymptotic distributions of 'psi-squared' goodness of fit criteria for *m*th order Markov chains," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1123–33.

[15] Leo A. Goodman, "A note on Stepanow's tests for Markov chains," *Teoriya Veroyatnostei i ee Primereniya* (The Theory of Probability and Its Applications), Vol. 4 (1959), to be published.

[16] P. G. Hoel, "A test for Markov chains," *Biometrika*, Vol. 41 (1954), pp. 430–3.

[17] G. M. Jenkins, "Contribution to the discussion on 'Regression analysis of binary sequences' by D. R. Cox," *J. Roy. Stat. Soc. Series B (Methodological)*, Vol. 20 (1958), pp. 238–9.

[18] V. E. Stepanow, "Some statistical tests for Markov chains," (in Russian), *Teoriya Veroyatnostei i ee Primereniya* (The Theory of Probability and Its Applications), Vol. 2 (1957), pp. 143–4.

[19] P. Whittle, "Some distribution and moment formulas for Markov chains," *J. Roy. Stat. Soc. Series B (Methodological)*, Vol. 17 (1955), pp. 235–42.

CONSENSUS OF SUBJECTIVE PROBABILITIES: THE PARI-MUTUEL METHOD

By Edmund Eisenberg and David Gale

The Rand Corporation and *Brown University*

A certain probability space is contemplated by a group of m individuals, each of whom endows it with his own subjective probability distribution. Suppose, now, that we wish to form a distribution which represents, in some sense, a consensus of those individual distributions. Various possibilities suggest themselves: the average, the convolution—but wait. There actually exists a popular institution which, theoretically at least, does perform just such an aggregation of personal probabilities. We refer to the pari-mutuel method of betting on horse races. In this system the final "track's odds" on a given horse are proportional to the amount bet on the horse. We shall here investigate the type of consensus given by this mechanism, which turns out to be quite different from any of the obvious aggregation schemes that might occur to one.

In formulating the pari-mutuel model we assume the m individuals involved are bettors, labeled $B_1, \cdots, B_m$, concerned with a race involving n horses, labeled $H_1, \cdots, H_n$. We assume further that each B_i, after careful study of the form sheets, the condition of the track, and other relevant material, has arrived at an estimate of the relative merits of each of the H_j's which he expresses in quantitative terms. Specifically, we are given an $m \times n$ *subjective probability matrix* $P = (p_{ij})$ where p_{ij} is the probability, in the opinion of B_i, that H_j will win the race.

Having determined his subjective probability distribution, B_i will now bet the amount b_i, a fixed positive number called B_i's *budget*, in a way which maximizes his *subjective expectation*. This means, of course, that B_i will not necessarily bet the whole amount b_i on that H_j for which p_{ij} is largest. In general, B_i will "bet the odds," that is, he will wait until the final track odds, or more conveniently, track probabilities, are announced. If these are $\pi_1, \cdots, \pi_n$, he will examine the ratios p_{ij}/π_j and distribute b_i among those H_j for which this ratio is a maximum. We shall refer to this course of action as B_i's *strategy*.

A technical difficulty is immediately apparent. We have already stated that the final track probabilities $\pi_1, \cdots, \pi_n$ are proportional to the amounts bet on $H_1, \cdots, H_n$, respectively (this is true whether or not the track retains a percentage). Thus, in practice, the π_j's are not known until each B_i has made his bet. On the other hand, B_i must know $\pi_1, \cdots, \pi_n$ before he can determine his bet. There is, therefore, a serious question as to whether there exist final track probabilities and individuals' bets compatible both with the bettors' strategies and the pari-mutuel principle. It is the purpose of this note to show that such probabilities and bets do exist and that the probabilities are in fact unique,

Received June 16, 1958.

thus giving a well-defined notion of consensus. Of course, the "influence" of B_i on the consensus will depend on his budget b_i, the case of equal influence being, by definition, that of equal budgets.

It will be convenient to choose the unit of money so that $\sum_{i=1}^{m} b_i = 1$. We shall also assume that each column of the matrix P contains at least one positive entry. If this were not so then, say, $p_{ij} = 0$ for all i and none of the B_i's would bet on H_j under any circumstances. We could then eliminate H_j from consideration entirely.

We shall now arithmetize the conditions which must be satisfied under the pari-mutuel system. Let β_{ij} be the amount which B_i bets on H_j. These must satisfy the *budget relation*.

$$\sum_{j=1}^{n} \beta_{ij} = b_i . \tag{1}$$

Next, the *pari-mutuel* condition requires that

$$\sum_{i=1}^{m} \beta_{ij} = \pi_j , \tag{2}$$

which is simply the statement that the final track probability π_j is proportional to the total amount bet on H_j. Equality holds here because of the normalization of the monetary unit. (We are using Greek letters to represent unknowns, Latin letters for the given constants of the problem.)

Finally, we must express the fact that each B_i is maximizing his expectation. The reader will easily verify that the condition is the following:

$$\text{if } \mu_i = \max_s \frac{p_{is}}{\pi_s} \text{ and } \beta_{ij} > 0, \text{ then } \mu_i = \frac{p_{ij}}{\pi_j}, \tag{3}$$

which states that B_i bets only on those H_j's for which his expectation is a maximum.

Nonnegative numbers π_j and β_{ij} which satisfy (1), (2) and (3) are called *equilibrium probabilities* and *bets*. Their existence can be proved by means of fixed-point theorems. We prefer, however, to prove existence in an elementary manner using a variational method which seems to be of interest in itself. We define a function ϕ and show that the variables which maximize it correspond to a solution of (1), (2) and (3).

The function ϕ has mn arguments ξ_{ij} and is defined by the rule:

$$\phi(\xi_{11}, \cdots, \xi_{mn}) = \sum_{i=1}^{m} b_i \log \sum_{j=1}^{n} p_{ij} \xi_{ij} ,$$

the variables ξ_{ij} being restricted to the domain D defined by:

$$\xi_{ij} \geqq 0, \qquad \text{for all } i, j, \tag{4}$$

$$\sum_{i=1}^{m} \xi_{ij} = 1, \qquad \text{for all } j. \tag{5}$$

We shall come back and discuss the meaning of the function ϕ after we have shown its relation to the pari-mutuel problem.

If we include minus infinity in the range of ϕ, then ϕ is continuous on the compact set D, hence attains a maximum at some point $(\bar{\xi}_{11}, \cdots, \bar{\xi}_{mn})$ of D. At this maximum the term $\sum_{j=1}^{n} p_{ij}\xi_{ij}$ is positive for every i (otherwise ϕ would be minus infinity, which is clearly not its maximum value). The partial derivatives of ϕ at the maximum are given by

$$\frac{\partial \phi}{\partial \bar{\xi}_{ij}} = \frac{b_i p_{ij}}{\sum_s p_{is} \bar{\xi}_{is}}.$$

We now assert:

EXISTENCE THEOREM. *A set of equilibrium probabilities π_j and bets β_{ij} are given by*

$$\pi_j = \max_i \frac{\partial \phi}{\partial \bar{\xi}_{ij}} = \max_i \frac{b_i p_{ij}}{\sum_s p_{is} \bar{\xi}_{is}} \tag{6}$$

$$\beta_{ij} = \bar{\xi}_{ij} \pi_j . \tag{7}$$

PROOF. We must show that the numbers π_j, β_{ij} satisfy (1), (2) and (3). The *pari-mutuel* condition (2) follows at once upon summing (7) on i and using condition (5) on the $\bar{\xi}_{ij}$.

The verification of (1) and (3) depends on the fact that

$$\text{if } \bar{\xi}_{ij} > 0, \text{ then } \pi_j = \frac{\partial \phi}{\partial \bar{\xi}_{ij}}. \tag{8}$$

To see this, suppose (8) is false and for some i, j we have $\bar{\xi}_{ij} > 0$ and $\pi_j > \partial\phi/\partial\bar{\xi}_{ij}$. By definition of π_j we have $\pi_j = \partial\phi/\partial\bar{\xi}_{kj} > \partial\phi/\partial\bar{\xi}_{ij}$ for some index k. Which means that by slightly decreasing $\bar{\xi}_{ij}$ and increasing $\bar{\xi}_{kj}$ by the same amount (which would not violate (4) or (5)), we could increase the value of ϕ, which is impossible since we are already at a maximum. Thus (8) is established.

We next verify condition (1). From (7) and (8) we have

$$\beta_{ij} = \bar{\xi}_{ij} \pi_j = \bar{\xi}_{ij} \frac{b_i p_{ij}}{\sum_s p_{is} \bar{\xi}_{is}}.$$

Summing the above on j,

$$\sum_j \beta_{ij} = b_i \frac{\sum_j p_{ij} \bar{\xi}_{ij}}{\sum_s p_{is} \bar{\xi}_{is}} = b_i .$$

Finally, we must prove (3). Since we assumed that none of the columns of the matrix P is identically zero, we know that each π_j is positive. Thus from (6) we have

$$\frac{p_{ij}}{\pi_j} \leqq \frac{\sum_s p_{is} \bar{\xi}_{is}}{b_i} \tag{9}$$

and μ_i, as defined in (3), is $(1/b_i)\sum_s p_{is}\bar{\xi}_{is}$. We see, then, from (7) and (8) that if β_{ij} is positive, $\bar{\xi}_{ij}$ is positive and hence $\mu_i = p_{ij}/\pi_j$, thus showing that (3) holds. (The fact that the π_j's sum to 1 is, of course, a consequence of (1) and (2) and the normalization of the b_i's.) Q.E.D.

The function ϕ can be interpreted as follows. From (7) we have $\sum_j p_{ij}\bar{\xi}_{ij} = \sum_j \beta_{ij}(p_{ij}/\pi_j)$, which is exactly the subjective expectation of B_i when he bets β_{ij} on H_j with track probabilities $\pi_1, \cdots, \pi_n$. Thus, at equilibrium the bettors, as a group, maximize a weighted sum of logarithms of subjective expectations, the weights being the bettors' budgets. As noted previously, equilibrium probabilities turn out to be unique, although equilibrium bets need not be unique. Furthermore, not every collection of β_{ij}'s, obtained by having each B_i act according to his strategy at equilibrium probabilities, need be equilibrium bets.

As a final result we show

UNIQUENESS THEOREM. *Equilibrium probabilities are unique.*

PROOF. Let $\pi_1, \cdots, \pi_n$ and $\bar{\pi}_1, \cdots, \bar{\pi}_n$ be equilibrium probabilities, let β_{ij}, $\bar{\beta}_{ij}$ be corresponding bets, and μ_i, $\bar{\mu}_i$ as defined in (3). Then for all i, j, k we have:

$$\beta_{ij}\mu_i\pi_j = \beta_{ij}p_{ij} \leqq \beta_{ij}\bar{\mu}_i\bar{\pi}_j$$

$$\bar{\beta}_{ik}\bar{\mu}_i\bar{\pi}_k = \bar{\beta}_{ik}p_{ik} \leqq \bar{\beta}_{ik}\mu_i\pi_k$$

whence, since μ_i, $\bar{\mu}_i$, π_j, π_k are positive, $\beta_{ij}\bar{\beta}_{ik}(\bar{\pi}_k/\pi_k) \leqq \beta_{ij}\bar{\beta}_{ik}\bar{\pi}_j/\pi_j$. Summing on j, k we get: $b_i\sum_k \bar{\beta}_{ik}(\bar{\pi}_k/\pi_k) \leqq b_i\sum_j \beta_{ij}\bar{\pi}_j/\pi_j$; dividing by b_i and summing on i: $\sum_k (\bar{\pi}_k\bar{\pi}_k/\pi_k) \leqq \sum_j \bar{\pi}_j = 1$.

Let $x_k = \bar{\pi}_k/\sqrt{\pi_k}$, $y_k = \sqrt{\pi_k}$. From the Cauchy-Schwartz inequality we have:

$$\left(\sum x_k y_k\right)^2 = \left(\sum \bar{\pi}_k\right)^2 = 1 \leqq \left(\sum x_k^2\right)\left(\sum y_k^2\right) = \left(\sum \frac{\bar{\pi}_k\bar{\pi}_k}{\pi_k}\right)\left(\sum \pi_k\right) \leqq 1.$$

Thus the vectors $(x_1, \cdots, x_n)$, $(y_1, \cdots, y_n)$ are dependent and $\bar{\pi}_k = \mu\pi_k$. But $\sum \bar{\pi}_k = \sum \pi_k = 1$, hence $\mu = 1$ and $\pi_k = \bar{\pi}_k$, proving uniqueness.

The referee has suggested the following instructive example which indicates the somewhat "pathological" nature of the pari-mutuel consensus. In the case of two bettors with equal budgets if the first bettor's subjective probability distribution on two horses is $(\frac{1}{2}, \frac{1}{2})$, then the equilibrium probabilities will be $(\frac{1}{2}, \frac{1}{2})$ regardless of the subjective probabilities of the second bettor, as the reader will easily verify.

A NOTE ON PERFECT PROBABILITY[1]

By Gopinath Kallianpur[2]

Michigan State University

1. Introduction. The purpose of this note is to define and characterize a class of perfect probability spaces which we shall call D-spaces. Gnedenko and Kolmogorov seem to have been the first to introduce explicitly the notion of perfect measure [1], although a special case ("normal space") was studied by Halmos and von Neumann as long ago as 1942 [2]. An illuminating appendix by Doob in [1] (see also his remarks in the appendix to his own book [3]) further testifies to the fact that the notion of perfectness of a measure has been well known to mathematicians for quite some time.

The triplet $(\Omega, \mathfrak{F}, \mu)$ is said to be a perfect probability space if μ is a probability over the σ algebra $\mathfrak{F}$ of subsets of Ω and if for every univalent, real valued, $\mathfrak{F}$-measurable function f the following is true: For every linear set A such that $f^{-1}(A) \in \mathfrak{F}$, there exists a linear Borel set B with $B \subseteq A$ and

$$\mu\{f^{-1}(B)\} = \mu\{f^{-1}(A)\}.$$

While a perfect probability space $(\Omega, \mathfrak{F}, \mu)$ has many desirable properties ([1], [4]), the definition of perfectness clearly involves the measure μ in an essential manner. This raises the interesting question of defining classes of measurable spaces $(\Omega, \mathfrak{F})$ with the property that for every probability μ, the space $(\Omega, \mathfrak{F}, \mu)$ is perfect. The Lusin spaces introduced by Blackwell [4] as well as the D-spaces to be defined in the next section, possess this property. Theorem 3 gives a necessary and sufficient criterion for a D-space. This result is similar to (though not identical with) an unsolved problem posed by Blackwell for Lusin spaces ([4] Problem 2).

2. D-spaces: definition and characterization. We shall say that a linear set A is a D-set if A is measurable with respect to F for every Lebesgue Stieltjes probability measure F. Borel sets and analytic sets are examples of D-sets.

A measurable space $(\Omega, \mathfrak{F})$ will be called a D-space if

(1) $\mathfrak{F}$ is a separable σ-field of subsets of Ω, and

(2) The range of every univalent, real valued, $\mathfrak{F}$-measurable function f is a D-set.

THEOREM 1. *Let $(\Omega, \mathfrak{F})$ be a D-space. Then if $\mathfrak{A} \subseteq \mathfrak{F}$ is any separable sub σ-field of $\mathfrak{F}$-sets the probability space $(\Omega, \mathfrak{A}, \mu)$ is perfect for every probability μ defined*

Received September 26, 1958.

[1] This research was supported in part by the United States Air Force through the Air Force Office of Scientific Research of the Air Research amd Development Command, under Contract No. AF 18 (600) - 442, Columbia University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] On leave from the Indian Statistical Institute.

over $\mathfrak{A}$. In particular, $(\Omega, \mathfrak{F}, \mu)$ is perfect for every probability μ defined over $\mathfrak{F}$. The proof of this theorem is based on the following

LEMMA. *A necessary and sufficient condition for $(\Omega, \mathfrak{F}, \mu)$ to be perfect is that for every univalent, real valued, $\mathfrak{F}$-measurable function f there exists an $\mathfrak{F}$-set Ω_0 such that*

$$\mu(\Omega_0) = 1 \text{ and } f(\Omega_0) \text{ is a Borel set.} \tag{2.1}$$

The above lemma is known in the literature, but we shall give a short proof of the sufficiency. The necessity of (2.1) is almost obvious and its proof is omitted. If (2.1) holds, Halmos and von Neumann have shown [2] that to every $\mathfrak{F}$-measurable function f and every $\mathfrak{F}$-set B corresponds a measurable set B_0 contained in B such that $\mu(B_0) = \mu(B)$ and $f(B_0)$ is a Borel set. Hence, if A is any linear set with $f^{-1}A \,\varepsilon\, \mathfrak{F}$, there exists a measurable subset X_0 of $f^{-1}A$ such that $\mu(X_0) = \mu(f^{-1}A)$ and $f(X_0)$ is a Borel set. Since $f(X_0) \subseteq A$ we may write $f^{-1}A$ in the form $f^{-1}A = f^{-1}\{f(X_0)\} \cup N$, where $N \subseteq f^{-1}(A) - X_0$ and $\mu(N) = 0$. Writing $B = f(X_0)$ we have B a Borel set, contained in A and $\mu(f^{-1}B) = \mu(f^{-1}A)$, so that $(\Omega, \mathfrak{F}, \mu)$ is perfect.

To prove Theorem 1, let μ be an arbitrary probability over $\mathfrak{F}$ and $\mathfrak{F}_\mu$ the σ-field obtained by completing $\mathfrak{F}$ with respect to μ. For any $\mathfrak{F}$-measurable f let $\mathfrak{F}^*$ be the σ-field of linear sets A such that $f^{-1}A \,\varepsilon\, \mathfrak{F}_\mu$, and let μ_f be the probability over $\mathfrak{F}^*$ defined by $\mu_f(A) = \mu(f^{-1}A)$. μ_f is then a complete probability over $\mathfrak{F}^*$. Finally, if F is the Lebesgue Stieltjes measure generated by the distribution function of f and $\mathfrak{A}_F$ the σ-field of F-measurable sets, it is easy to see that $\mathfrak{A}_F \subseteq \mathfrak{F}^*$ and $\mu_f = F$ on $\mathfrak{A}_F$. Since $(\Omega, \mathfrak{F})$ is a D-space by our assumption, $f(\Omega) \,\varepsilon\, \mathfrak{A}_F$, so that there exists a Borel set $B \subseteq f(\Omega)$ such that $F(B) = 1$. Since F and μ_f agree on Borel sets, $\mu_f(B) = 1$. Now setting $\Omega_0 = f^{-1}(B)$ we have $\mu(\Omega_0) = 1$ and $f(\Omega_0) = B$, a Borel set. The perfectness of $(\Omega, \mathfrak{F}, \mu)$ then follows by the Lemma. Since $(\Omega, \mathfrak{F})$ is a D-space and every $\mathfrak{A}$-measurable f is *a-fortiori* $\mathfrak{F}$-measurable, $(\Omega, \mathfrak{A})$ is a D-space. The perfectness of $(\Omega, \mathfrak{A}, \mu)$ for every μ follows on replacing $\mathfrak{F}$ by $\mathfrak{A}$ in the above proof.

The converse of Theorem 1 is given by

THEOREM 2. *Let $(\Omega, \mathfrak{F})$ be a measurable space with the following property:*

(I) If $\mathfrak{A}$ is any separable sub σ-field of $\mathfrak{F}$-sets, the probability space $(\Omega, \mathfrak{A}, \mu)$ is perfect for every probability μ defined over $\mathfrak{A}$.

Then, the range of every $\mathfrak{F}$-measurable function f is a D-set.

PROOF OF THEOREM 2. Let $\mathfrak{A}$ be a separable sub σ-field of $\mathfrak{F}$. Then there exists an $\mathfrak{F}$-measurable function f such that $\mathfrak{A}$ is the minimal σ-field with respect to which f is measurable. In other words, $\mathfrak{A}$ is the σ-field of sets $f^{-1}(E)$ where E is a Borel set. Let ν be a Lebesgue Stieltjes probability measure and $\mathfrak{S}_\nu$ the σ-field of ν-measurable sets. If E is any subset of the real line it is known that there exists a Borel set F such that $F \supset E$ and such that, for every Borel set $B \subset F - E$, we have $\nu(B) = 0$. We also have $\nu^*(E) = \nu(F)$, ν^* being outer ν-measure ([5], pp. 50–51). Such a set F we shall call a *ν-cover* of E. Let $R_f = f(\Omega)$, the range of f. Denote by K_1 the ν-cover of R_f. If $\nu^*(R_f) = 0$, it is a well-known fact that $R_f \,\varepsilon\, \mathfrak{S}_\nu$. If $\nu^*(R_f) > 0$ we also have $\nu(K_1) > 0$ and we may now define

a probability μ on $\mathfrak{A}$ as follows: If $A \in \mathfrak{A}$ then $A = f^{-1}(E)$ for some Borel set E. Define $(1)\mu(A) = \nu(E \cap K_1)/\nu(K_1)$. First we show that (1) defines μ uniquely. Suppose E_1 and E_2 are two Borel sets such that $A = f^{-1}(E_1) = f^{-1}(E_2)$. Then clearly the Borel sets $E_1 - (E_1 \cap E_2)$ and $E_2 - (E_1 \cap E_2)$ are contained in the complement of R_f. For $i = 1, 2$, $(E_i - E_1 \cap E_2) \cap K_1 \subset K_1 - R_f$ and $(E_i - E_1 \cap E_2) \cap K_1$ is a Borel set. Since K_1 is a ν-cover of R_f we have $\nu[(E_i - E_1 \cap E_2) \cap K_1] = 0$. From this it follows that $\nu(E_1 \cap K_1) = \nu(E_2 \cap K_1)$, proving that $\mu(A)$ is uniquely defined. Since $\Omega = f^{-1}R_1$, R_1 being the real line, we have according to $(1)\mu(\Omega) = 1$. Thus μ is a probability defined over sets of $\mathfrak{A}$. However, it is to be remembered that μ is not defined for all sets of $\mathfrak{F}$.

By the hypothesis of the theorem, $(\Omega, \mathfrak{A}, \mu)$ is perfect. Therefore, since $f^{-1}(R_f) \in \mathfrak{A}$, there exists a Borel set $K_0 \subset R_f$ such that

$$\mu(f^{-1}K_0) = \mu(f^{-1}R_f) = \mu(\Omega) = 1.$$

But by the definition of μ,

$$\mu(f^{-1}K_0) = \frac{\nu(K_0 \cap K_1)}{\nu(K_1)} = \frac{\nu(K_0)}{\nu(K_1)}, \quad \text{so that} \quad \nu(K_0) = \nu(K_1).$$

Thus, there exist two Borel sets K_0 and K_1 such that $K_0 \subset R_f \subset K_1$ and $\nu(K_0) = \nu(K_1)$. Hence, remembering that ν is complete, we have $R_f \in \mathfrak{F}_\nu$. Since ν is an arbitrary Lebesgue Stieltjes measure, the theorem is proved.

From the two theorems proved above we obtain the following characterization of a D-space ($\mathfrak{F}$ is assumed to be separable):

THEOREM 3. *A necessary and sufficient condition for a measurable space* $(\Omega, \mathfrak{F})$ *to be a* D*-space is that condition* (I) *of Theorem* 2 *be satisfied.*

If f is any $\mathfrak{F}$-measurable function, $\mathfrak{A}_f$ the minimal σ-field with respect to which f is measurable is known to be separable. Hence, Theorem 3 can also be given the following form:

THEOREM 3′. *Let* $(\Omega, \mathfrak{F})$ *be a measurable space,* f *any univalent, real valued* $\mathfrak{F}$*-measurable function and* $\mathfrak{A}_f$ *the* σ*-field defined as above. Then, a necessary and sufficient condition in order that* $(\Omega, \mathfrak{A}_f, \mu)$ *be perfect for all probability measures* μ *is that* $(\Omega, \mathfrak{F})$ *be a* D*-space.*

Recently Blackwell has defined a Lusin space to be any $(\Omega, \mathfrak{F})$ with $\mathfrak{F}$, a separable σ-field and with the property that the range of every real valued $\mathfrak{F}$-measurable f is an analytic set. Since analytic sets in metric spaces are Lebesgue Stieltjes measurable for every Lebesgue Stieltjes measure, it follows that a Lusin space is also a D-space. Whether, in reality, the concept of a D-space is more general than that of a Lusin space, we do not know. We have not succeeded in demonstrating the existence of a $(\Omega, \mathfrak{F})$ and a real-valued $\mathfrak{F}$-measurable function whose range is a D-set other than an analytic or a Borel set. As far as we are able to determine, very little seems to be known about the properties of D-sets beyond the fact that a set S on the real line is a D-set if and only if every homeomorphic image of S situated on the real line is Lebesgue measurable [6]. Nevertheless, the introduction of the notion of D-space is justified by the fact

that we are able to prove a characterizing property of such spaces given by Theorem 3, whereas we are unable to prove a similar result for Lusin spaces. In fact, Blackwell has posed the following unsolved problem for Lusin spaces: If $(\Omega, \mathfrak{F})$, with $\mathfrak{F}$ separable, is such that $(\Omega, \mathfrak{F}, \mu)$ is perfect for every probability μ defined on $\mathfrak{F}$, is $(\Omega, \mathfrak{F})$ a Lusin space? Theorem 2 proves a somewhat weaker property for D-spaces. Condition (I) of Theorem 2 is more stringent than the restriction that $(\Omega, \mathfrak{F}, \mu)$ be perfect for every probability μ on $\mathfrak{F}$. If the latter is given it is, of course, true that $(\Omega, \mathfrak{A}, \mu)$ is perfect, $\mathfrak{A}$ being any sub σ-field of $\mathfrak{F}$ and μ being regarded as the contraction on $\mathfrak{A}$ of the probability μ already defined on $\mathfrak{F}$. Condition (I) goes beyond this in requiring the perfectness of $(\Omega, \mathfrak{A}, \mu)$, ($\mathfrak{A}$ an arbitrary, separable sub σ-field of $\mathfrak{F}$) for all probabilities μ on $\mathfrak{A}$ and not merely for those μ which are contractions of probabilities defined over the larger σ-field $\mathfrak{F}$.

REFERENCES

[1] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, translated by K. L. Chung with an appendix by J. L. Doob, Cambridge, Addison-Wesley, 1954.

[2] P. R. Halmos and J. v. Neumann, "Operator methods in classical mechanics II," *Ann. of Math.*, Vol. 43 (1942), pp. 332–50.

[3] J. L. Doob, *Stochastic Processes*, New York, John Wiley and Sons, 1953.

[4] David Blackwell, "On a class of probability spaces," *Third Berkeley Symposium*, Vol. 2 (1956), pp. 1–6.

[5] Paul R. Halmos, *Measure Theory*, New York, D. Van Nostrand, 1950.

[6] E. Marczewski, "Remarque sur la mesurabilite absolue" (Abstract), *Colloquium Mathematicum*, Vol. 1 (1948), pp. 42–43.

ON THE DISTRIBUTION OF THE KOLMOGOROV-SMIRNOV D-STATISTIC

By Pedro Egydio de Oliveira Carvalho[1]

University of São Paulo

Summary. Gnedenko and Korolyuk [1] have pointed out that the exact distribution of the Kolmogorov-Smirnov D-statistic can be obtained explicitly by solving a certain double-boundary random walk problem, which, in turn, is solved by the principle of reflection. This principle is employed here in what is believed to be a new way to derive Gnedenko's and Korolyuk's result.

A random walk problem. Let us consider a random walk on the half plane $(t > 0, s)$, starting from the origin, such that at every point (t, s) there are two possible steps to take, either to $(t + 1, s + 1)$ or to $(t + 1, s - 1)$, each with equal probability and for some positive integer n, consider the paths from the origin to the point $(2n, 0)$. Among these, let us denote for any non-negative integer $k \leqq n$, the set of all paths that have a point on the line $s = k$ by $C(d_y \geqq k)$, the set of all those paths that reach the line $s = -k$ by $C(d_x \geqq k)$, the set of all those that have a point on at least one of these two lines by $C(d \geqq k)$, and the set of all those that reach both $s = \alpha k$ and $s = -k$, but go to the $s = \alpha k$ line first, by $C(d_y \geqq \alpha k \to d_x \geqq k)$. Let the number of elements in $C(\ )$ be $C^*(\ )$.

While it is well known (p. 70, 2) that

$$C^*(d_x \geqq k) = \binom{2n}{n+k} \tag{1}$$

$C^*(d \geqq k)$ is more difficult to calculate.

Clearly, $C^*(d \geqq k) = C^*(d_x \geqq k) + C^*(d_y \geqq k)$ less the number of paths in $C(d_x \geqq k) \cap C(d_y \geqq k)$ or

$$\begin{aligned} C^*(d \geqq k) = {} & C^*(d_x \geqq k) + C^*(d_y \geqq k) - C^*(d_x \geqq k \to d_y \geqq k) \\ & - C^*(d_y \geqq k \to d_x \geqq k) \end{aligned} \tag{2}$$

and by symmetry

$$C^*(d \geqq k) = 2C^*(d_x \geqq k) - 2C^*(d_x \geqq k \to d_y \geqq k). \tag{3}$$

Because of (1) it remains to calculate the last term in (3). As the first step, we show that for $i = 2, 3, \cdots, [n/k]$,

$$C^*(d_x \geqq ik) - C^*(d_x \geqq ik \to d_y \geqq k) = C^*(d_x \geqq (i-1)k \to d_y \geqq k). \tag{4}$$

Received January 20, 1958; revised June 12, 1958.

[1] Posthumous note. Revisions were made and references to literature supplemented following the referee's suggestions by Agnes Berger and Ruth Gold, School of Public Health and Administrative Medicine, Columbia University.

A path counted on the left side of (4) is one of two types. The first type reaches $s = -ik$, but does not reach $s = k$, the second reaches both lines but reaches $s = k$ first.

Let P be a path of the first type. By definition, it has points on $s = -ik$. Let p be the first of these. There must be points of P on $s = -(i-1)k$ to the left and also to the right of p; let the closest one to the left be p_{1l} and to the right, p_{1r}. Replace the portion of p from p_{1l} to p_{1r} by its reflection about $s = -(i-1)k$. The new path P' contains the image of p, say p'_1, falling on the line $s = -(i-2)k$. On $s = -(i-2)k$, let the points of P' nearest to p_{1r} be p'_{2l} on the left and p_{2r} on the right. Reflect P' between p'_{2l} and p_{2r} about $s = -(i-2)k$, to get a new path P''. On $s = -(i-3)k$, let the points of P'' nearest to p_{2r} be p''_{3l} on the left and p_{3r} on the right. Continuing in this manner, i reflections will lead to a path $P^{(i)}$ that goes first to $s = -(i-1)k$ and then to $s = k$, but does not reach $s = -ik$ except possibly after reaching $s = k$. Thus

$$P^{(i)} \varepsilon C(d_x \geqq (i-1)k \rightarrow d_y \geqq k)$$

but

$$P^{(i)} \varepsilon C(d_x \geqq ik \rightarrow d_y \geqq k).$$

Conversely, let Q be any path such that

$$Q \varepsilon C(d_x \geqq (i-1)k \rightarrow d_y \geqq k)$$

but

$$Q \varepsilon C(d_x \geqq ik \rightarrow d_y \geqq k).$$

Q has points on $s = k$, let q be the first of these. On $s = 0$, let q_l and q_r be the nearest points of Q to the left and right of q, respectively. Let all the other points of Q on $s = 0$ to the right of q_r be $a_1, \cdots, a_m$, in order. Let us reflect the portions of Q between q_l and q_r and at the same time between all those points a_i, a_{i+1} between which Q reaches $s = k$ about the line $s = 0$. The new path Q' does not reach k. The reflection of q, say q', lies on $s = -k$. Next reflect Q' between q' and the nearest point to the left of it on $s = -k$. Continuing in the same manner, the ith reflection will produce a path that reaches $s = -ik$ but never reaches $s = k$.

Let U be a path of the second type, i.e., one that reaches $s = k$ first and then reaches $s = -ik$. Let $p = (t, 0)$ be the first return of U to $s = 0$ after having reached $s = -ik$. Let the portion of U between $(0, 0)$ and $(t, 0)$ be represented by the ordered sequence $\epsilon_1, \epsilon_2, \cdots, \epsilon_t$, where ϵ_j is a vector of length $\sqrt{2}$ and slope $+1$ or -1. Let U' be a path such that from $(0, 0)$ to $(t, 0)$ it is given by the reversed sequence $\epsilon_t, \epsilon_{t-1}, \cdots, \epsilon_1$ and coincides with U from $(t, 0)$ to $(2n, 0)$. U' is clearly in $C(d_x \geqq ik \rightarrow d_y \geqq k)$ and therefore also in $C(d_x \geqq (i-1)k \rightarrow d_y \geqq k)$.

Conversely, let V be a path such that

$$V \varepsilon C(d_x \geqq (i-1)k \rightarrow d_y \geqq k)$$

and

$$V \varepsilon C(d_s \geqq ik \to d_v \geqq k)$$

and let $q_0' = (t', 0)$ be the first return of V to $s = 0$ after having reached $s = k$. Reversing the steps between $(0, 0)$ and q_0' uniquely determines a path of type II, completing the proof of (4).

Note that (4) has the structure

$$A_i - B_i = B_{i-1}$$

where $A_i = C^*(d_s \geqq ik)$ is known by (1). Thus knowing B_i for any i implies knowing all B_j for $j < i$. But for $i = [n/k]$, (4) gives

$$\binom{2n}{n + \left[\frac{n}{k}\right]} - 0 = C^*\left(d_s \geqq \left(\left[\frac{n}{k}\right] - 1\right)k \to d_v \geqq k\right) = 1 = B_{[n/k]-1}.$$

Carrying out the substitutions gives

$$C^*(d_s \geqq k \to d_v \geqq k) = \sum_{i=2}^{[n/k]} \binom{2n}{n + ik} (-1)^i$$

and from (3)

$$C^*(d \geqq k) = 2 \sum_{i=1}^{[n/k]} \binom{2n}{n + ik} (-1)^{i+1}$$

Application to the Kolmogorov-Smirnov problem. Let $X = (x_1 < x_2 < \cdots < x_n)$ and $Y = (y_1 < y_2 < \cdots < y_n)$ be two independent samples of ordered independent observations having the same continuous cumulative distribution function. Suppose $x_i \neq y_j$, $(i, j = 1, 2, \cdots, n)$ and let the two samples be combined and arranged in increasing order of magnitude, say $Z = (z_1 < z_2 < \cdots < z_{2n})$. Let $S_n(x)$ be the number of observed values x_i which are less than or equal to x and $S_n'(x)$ the number of observed y_j's less than or equal to x.

Let

$$D^+ = \max_x \left(S_n(x) - S_n'(x)\right)$$

and

$$D = \max_x \left| S_n(x) - S_n'(x) \right|.$$

The limiting distribution of D was found by Kolmogorov [3, 4] and Smirnov [5, 6, see also 7] and an iterative method for its exact distribution has been given by Massey [8]. Gnedenko and Korolyuk recognized that a one to one correspondence exists between the set of all Z and all paths from $(0, 0)$ to $(2n, 0)$ in the above discussed random walk: Starting from $(0, 0)$, we move to $(1, 1)$ if z_1 is from Y, to $(1, -1)$, if z_1 is from X and so on. In particular, samples Z for

which $D \geqq k$ correspond to paths in $C(d \geqq k)$ and vice versa. Thus we get Gnedenko's and Korolyuk's result

$$P\{D < k\} = 1 - \frac{C^*(d \geqq k)}{\binom{2n}{n}}.$$

REFERENCES

[1] B. V. Gnedenko and V. S. Korolyuk, "On the maximum discrepancy between two empirical distributions". *Doklady Akad. Nauk. SSSR (N.S.)*, Vol. 80 (1951), pp. 525–528. Reviewed by W. Feller in *Mathematical Reviews*, Vol. 13 (1952), pp. 570–571.

[2] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 1, John Wiley and Sons, Inc., New York, 1957.

[3] A. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione", *Inst. Ital. Attuari, Giorn.*, Vol. 4 (1933), pp. 1–11.

[4] A. Kolmogorov, "Über die Grenzwertsätze der Wahrscheinlichkeitsrechnung", *Bulletin [Izvestija] Academie des Sciences URSS*, (1933), pp. 363–372.

[5] N. Smirnov, "Ob uklonenijah empiričeskoi krivoi raspredelenija", *Recueil Mathematique (Matematičeskii Sbornik)*, N.S. Vol. 6 (48) (1939), pp. 3–26.

[6] N. Smirnov, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples", *Bulletin Mathématique de l'Université de Moscou*, Vol. 2 (1939), fasc. 2.

[7] W. Feller, "On the Kolmogorov-Smirnov limit theorems for empirical distributions", *Ann. Math. Stat.*, Vol. XIX, No. 2 (1948), pp. 177–189.

[8] F. Massey, Jr., "The distribution of the maximum deviation between two sample cumulative step functions", *Ann. Math. Stat.*, Vol. 22, No. 1 (1951), pp. 125–131.

EQUALITY OF MORE THAN TWO VARIANCES AND OF MORE THAN TWO DISPERSION MATRICES AGAINST CERTAIN ALTERNATIVES[1]

By R. Gnanadesikan[2]

University of North Carolina

0. Introduction and summary. In this paper, using the heuristic union-intersection principle [4], two tests are proposed, and the associated simultaneous confidence bounds on parametric functions which are measures of a certain type of departure from the respective null hypotheses are obtained. The first test is for the equality of $(k+1)$ variances $(k \geqq 2)$ of $(k+1)$ univariate normal populations, wherein we choose one of the variances as a standard (of course, unknown), and compare the other k variances with it. The alternative to the hypothesis is that not all the k variances are equal to the standard one. The proposed test may be called the simultaneous variance ratios test. The well-known Hartley's $F_{\max}$ test [2] for the case of equal sample sizes is not equivalent to the present test even when all samples are of the same size since the alternatives in the two cases are different. In the alternative in Hartley's test, aside from the inequality of the k variances to the standard one, the mutual inequality of the k variances also plays an important role. The second test proposed in this paper, is a multivariate extension of the first. This paper also considers the distribution problems that arise in connection with both the tests. The non-availability of tables at the moment makes the immediate practical application of the tests and the associated confidence bounds not possible.

Sections 1, 2, and 3 deal with the univariate problem and Sections 4 and 5 deal with the multivariate extensions.

1. The simultaneous variance ratios test. For $(k+1)$ univariate normal populations we want to test the composite hypothesis $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma_0^2$. Suppose we have independent random samples of sizes $(n_i + 1)$, $i = 0, 1, 2, \cdots, k$, from the $(k+1)$ populations, and let s_i^2 be the estimate of σ_i^2 based on n_i degrees of freedom for $i = 0, 1, \cdots, k$. Let us choose σ_0^2 as standard and compare $\sigma_1^2, \cdots, \sigma_k^2$ with σ_0^2, so that H_0 is equivalent to $H_0' : \sigma_1^2/\sigma_0^2 = \cdots = \sigma_k^2/\sigma_0^2 = 1$. The alternative hypothesis is H_1' : Not H_0', i.e., at least one $\sigma_i^2/\sigma_0^2 \neq 1$. For each hypothesis like $H_{0i}' : \sigma_i^2/\sigma_0^2 = 1$ against $H_{1i}' : \sigma_i^2/\sigma_0^2 \neq 1$, we have the well-known test with the acceptance region,

$$F_{i1} \leqq F_i(n_i, n_0) \leqq F_{i2}, \tag{1.1}$$

where $F_i(n_i, n_0) = s_i^2/s_0^2$ has the central F-distribution with n_i and n_0 degrees of

Received August 5, 1957; revised July 29, 1958.

[1] This research was sponsored by the United States Air Force through the Office of the Air Research and Development Command.

[2] Now with the Procter and Gamble Company, Cincinnati, Ohio.

freedom under H'_{0i}. It is also easily seen that $H'_0 = \bigcap_{i=1,\cdots,k} H'_{0i}$ and $H'_1 = \bigcup_{i=1,\cdots,k} H'_{1i}$. Therefore, by the heuristic union-intersection principle [4], we shall take for our test of H'_0, i.e., of H_0, the acceptance region,

$$F_{11} \leqq F_1(n_1, n_0) \leqq F_{12}, F_{21} \leqq F_2(n_2, n_0) \leqq F_{22}, \cdots, F_{k1} \leqq F_k(n_k, n_0) \leqq F_{k2}, \tag{1.2}$$

which is the intersection (over i) of the regions (1.1). For the critical region, therefore, we take the union (over i) of the complements of the regions (1.1).

The optimum choice of F_{i1}, F_{i2}, for $i = 1, 2, \cdots, k$, is not known, and, in the absence of this knowledge, the following choice is suggested as one possible way:

Following the usual procedure for obtaining a Type I union-intersection region, choose F_{i1} and F_{i2} such that all the individual regions (1.1) will have the same size $(1 - \alpha^*)$, where α^* is such that the size of the intersection (1.2) is $(1 - \alpha)$, for a preassigned α. In general, of course, $(1 - \alpha) \neq (1 - \alpha^*)^k$, but assuming non-triviality, given α we can determine α^* and vice versa. This condition, however, still does not determine the region (1.2) completely. In order to do so, we impose the further condition that, for each i, the test with acceptance region (1.1) be locally unbiased (in the sense of Neyman). This latter condition, as will be shown in Section 2, ensures a desirable property of the simultaneous variance ratios test with acceptance region (1.2).

$F_i(n_i, n_0)$, for $i = 1, 2, \cdots, k$, are quasi-independent variance ratios in the sense of [3], i.e., the numerators, except for constant multipliers, are distributed independently as χ^2 variates and are so distributed independently of their common denominator which also, except for a constant multiplier, is distributed as a χ^2 variate. The joint distribution of such quasi-independent variance ratios is given in [3] and using an approach which is essentially the same as that contained in that paper we can obtain a recurrence relation to aid us in evaluating the probability integral associated with the Simultaneous Variance Ratios test (1.2). It must be noted, however, that the recurrence relation solves the problem only in theory and for practical purposes tables of the probability integral need to be constructed.

2. Properties of the power of the test proposed in Section 1. We shall first note that the power, or, equivalently, the probability of the second kind of error, β, of the test could involve as parameters only the k ratios $\delta_i = \sigma_i^2/\sigma_0^2$, $(i = 1, 2, \cdots, k)$.

$$\beta = P[F_{11} \leqq F_1(n_1, n_0) \leqq F_{12}, \cdots, F_{k1} \leqq F_k(n_k, n_0) \leqq F_{k2} \mid H'_1]$$

$$= P\left[\frac{F_{11}}{\delta_1} \leqq \frac{F_1(n_1, n_0)}{\delta_1} \leqq \frac{F_{12}}{\delta_1}, \cdots, \frac{F_{k1}}{\delta_k} \leqq \frac{F_k(n_k, n_0)}{\delta_k} \leqq \frac{F_{k2}}{\delta_k} \,\middle|\, H'_0\right],$$

where for $i = 1, 2, \cdots, k$, $F_i(n_i, n_0)/\delta_i$ has a central F-distribution with degrees of freedom n_i and n_0 and the different F's are quasi-independent. It now follows that β could involve as parameters only $\delta_1, \delta_2, \cdots, \delta_k$.

We shall next show that, for the choice of F_{i1}, F_{i2} mentioned in Section 1, the power of the test has the monotonicity property, i.e., that as each δ_i, $(i = 1, 2, \cdots, k)$, tends away from unity the power monotonically increases. It is to be noted that Ramachandran [3] has proved a similar property of the simultaneous analysis of variance test.

Let us write $v_i = s_i^2/\sigma_i^2$, $i = 0, 1, 2, \cdots, k$, and let $p(v_i)$ denote the probability density function of χ^2 with n_i degrees of freedom. Then we have

$$\frac{\partial \beta}{\partial \delta_1} = \frac{\partial}{\partial \delta_1} \int_0^\infty p(v_0) \left[\prod_{i=1}^{k} \int_{\frac{F_{i1}v_0}{\delta_i}}^{\frac{F_{i2}v_0}{\delta_i}} p(v_i)\, dv_i \right] dv_0$$

(2.1)

$$= \int_0^\infty p(v_0)\, dv_0 \frac{\partial}{\partial \delta_1} \left[\prod_{i=1}^{k} \int_{\frac{F_{i1}v_0}{\delta_i}}^{\frac{F_{i2}v_0}{\delta_i}} p(v_i)\, dv_i \right],$$

this being valid,

$$= \int_0^\infty p(v_0) \left[\left(\frac{\partial}{\partial \delta_1} \int_{\frac{F_{11}v_0}{\delta_1}}^{\frac{F_{12}v_0}{\delta_1}} p(v_1)\, dv_1 \right) \times \prod_{i=2}^{k} \int_{\frac{F_{i1}v_0}{\delta_i}}^{\frac{F_{i2}v_0}{\delta_i}} p(v_i)\, dv_i \right] dv_0$$

$$= \int_0^\infty p(v_0) \left[\left(\frac{n_1}{2} \right)^{\frac{n_1}{2}} \cdot \frac{1}{\Gamma\left(\frac{n_1}{2} \right)} \cdot \frac{1}{\delta_1} \left\{ \left(\frac{F_{11} v_0}{\delta_1} \right)^{\frac{n_1}{2}} e^{-\frac{n_1 F_{11} v_0}{2\delta_1}} \right.\right.$$

$$\left.\left. - \left(\frac{F_{12} v_0}{\delta_1} \right)^{\frac{n_1}{2}} e^{-\frac{n_1 F_{12} v_0}{2\delta_1}} \right\} \prod_{i=2}^{k} \int_{\frac{F_{i1}v_0}{\delta_i}}^{\frac{F_{i2}v_0}{\delta_i}} p(v_i)\, dv_i \right] dv_0$$

$$= \text{const.} \int_0^\infty p(v_0) f(v_0)\, dv_0,$$

say, where

$$f(v_0) = v_0^{\frac{n_1}{2}} \left[F_{11}^{\frac{n_1}{2}} e^{-\frac{n_1 F_{11} v_0}{2\delta_1}} - F_{12}^{\frac{n_1}{2}} e^{-\frac{n_1 F_{12} v_0}{2\delta_1}} \right]$$

$$\times \prod_{i=2}^{k} \int_{\frac{F_{i1}v_0}{\delta_i}}^{\frac{F_{i2}v_0}{\delta_i}} p(v_i)\, dv_i,$$

and the constant factor is non-negative. Noticing that $v_0, \delta_i, F_{i2} > F_{i1}$, $(i = 1, 2, \cdots, k)$, are all essentially positive and that each of the integrals in the product $\prod_{i=2}^{k} \cdots$ is positive (lying between 0 and 1), we may apply a well-known result in Calculus to obtain, $\partial\beta/\partial\delta_1 \gtrless 0$ according as

$$F_{11}^{\frac{n_1}{2}} e^{-\frac{n_1 F_{11} v_0}{2\delta_1}} - F_{12}^{\frac{n_1}{2}} e^{-\frac{n_1 F_{12} v_0}{2\delta_1}} \gtrless 0.$$

i.e., according as

$$\left(\frac{F_{11}}{F_{12}} \right)^{\frac{n_1}{2}} \gtrless e^{-\frac{n_1 v_0}{2\delta_1}(F_{12} - F_{11})}. \tag{2.2}$$

It can be shown that the condition of local unbiasedness of the region, $F_{11} \leqq F_1(n_1, n_0) \leqq F_{12}$, reduces to,

$$\left(\frac{F_{11}}{F_{12}}\right)^{\frac{n_1}{2}} = e^{-\frac{n_1 v_0}{2}(F_{12}-F_{11})}. \tag{2.3}$$

Substituting in (2.2), we see, after some simplification, that

$$\partial\beta/\partial\delta_1 \gtrless 0 \text{ according as } \delta_1 \lesseqgtr 1 \text{ (irrespective of } v_0 > 0). \tag{2.4}$$

Hence, the power is a monotone increasing or decreasing function of δ_1 according as $\delta_1 \gtrless 1$. The same property with respect to $\delta_2, \cdots, \delta_k$ can be proved similarly.

Also, if $\delta'(1 \times k) = (\delta_1, \delta_2, \cdots, \delta_k)$, then from (2.1) and (2.3) it appears that if F_{i1}, F_{i2} are chosen so as to make the region (1.1) locally unbiased, then

$$\left.\frac{\partial\beta}{\partial\delta_i}\right|_{\delta'=(1,\cdots,1)} = 0. \tag{2.5}$$

Therefore, the proposed test is locally unbiased, and, as a consequence of its monotonicity property, it will be completely unbiased.

3. The associated simultaneous confidence bounds on σ_i^2/σ_0^2, $(i = 1, 2, \cdots, k)$. Under the alternative hypothesis it is known that $F_i(n_i, n_0)/\delta_i$, for $i = 1, 2, \cdots, k$, are distributed as quasi-independent variance ratios. Hence, we can make the following simultaneous statements:

$$F_{11} \leqq \frac{F_1(n_1, n_0)}{\delta_1} \leqq F_{12}, \cdots F_{k1} \leqq \frac{F_k(n_k, n_0)}{\delta_k} \leqq F_{k2}, \tag{3.1}$$

where F_{i1}, F_{i2} $(i = 1, 2, \cdots, k)$ are such that

$$P[F_{11} \leqq F_1(n_1, n_0) \leqq F_{12}, \cdots, F_{k1} \leqq F_k(n_k, n_0) \leqq F_{k2}] = (1 - \alpha), \tag{3.2}$$

so that the probability associated with (3.1) is $(1 - \alpha)$.

By inverting the statements (3.1), it is easily seen that we obtain the following simultaneous confidence interval statements,

$$\frac{s_i^2}{s_0^2 F_{i2}} \leqq \delta_i \leqq \frac{s_i^2}{s_0^2 F_{i1}}, \qquad i = 1, 2, \cdots, k, \tag{3.3}$$

with a joint confidence coefficient $(1 - \alpha)$.

These results are valid for all choices of F_{i1}, F_{i2}, $(i = 1, 2, \cdots, k)$, satisfying (3.2). However, if F_{i1}, F_{i2}, $(i = 1, 2, \cdots, k)$, are chosen as in Section 1, then, from the unbiasedness and monotonicity properties of the associated test, (proved in Section 2), we shall have the desirable property of monotonically increasing *shortness* (in terms of probability of covering wrong values) for the confidence bounds (3.3).

4. The multivariate test. The notation used in this and the following section is fairly standard and, for example, is the same as that used in [6], [7], [8]. The

case $k = 1$ is treated in these papers but now we shall proceed to consider the case $k \geqq 1$, and, as will be seen, the results for the case $k \geqq 1$ are similar in form to those for the case $k = 1$.

In the multivariate situation we need a test for the hypothesis of equality of the dispersion matrices of $(k + 1)$ non-singular p-variate normal populations, $N[\xi_i, \Sigma_i]$, $i = 0, 1, 2, \cdots, k$. That is, the null hypothesis is $H_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_k = \Sigma_0$. Suppose that $X_i[p \times (n_i + 1)]$, $i = 0, 1, \cdots, k$, where $p \leqq n_i$ for $i = 0, 1, \cdots, k$, are mutually independent random samples respectively from the $(k + 1)$ normal populations. Let

$$n_i S_i (p \times p) = X_i X_i' - (n_i + 1)\bar{x}_i \bar{x}_i', \quad i = 0, 1, 2, \cdots, k, \tag{4.1}$$

where $\bar{x}_i (p \times 1)$'s are the sample mean vectors and $S_0, S_1, \cdots, S_k$ are sample dispersion matrices estimating $\Sigma_0, \Sigma_1, \cdots, \Sigma_k$ respectively. The sample dispersion matrices have independent Wishart distributions with S_i having the distribution,

$$p(S_i)\, dS_i \propto |\Sigma_i|^{\frac{-n_i}{2}} \exp\left[-\frac{n_i}{2} \operatorname{tr} \Sigma_i^{-1} S_i\right] |S_i|^{\frac{n_i - p - 1}{2}}\, dS_i, \tag{4.2}$$

for $i = 0, 1, 2, \cdots, k$.

Just as in the univariate case discussed in Section 1, we may, for the multivariate case, choose Σ_0 as standard and compare the k matrices $\Sigma_1, \cdots, \Sigma_k$ with Σ_0. Notice that $S_0, S_1, \cdots, S_k$ are symmetric and almost everywhere (i.e., except on a set of probability measure zero) positive definite, and $\Sigma_0, \Sigma_1, \cdots, \Sigma_k$ are symmetric positive definite matrices being dispersion matrices of non-singular p-variate normal distributions.

Consider, in analogy with (1.2), the test for $H_0 : \Sigma_1 = \cdots = \Sigma_k = \Sigma_0$, whose acceptance region is

$$\lambda_{j1} \leqq \frac{c_{\min}(S_j)}{c_{\max}(S_0)} \leqq \frac{c_{\max}(S_j)}{c_{\min}(S_0)} \leqq \lambda_{j2}, \qquad j = 1, 2, \cdots, k. \tag{4.3}$$

Here also the optimum choice of $\lambda_{j1}, \lambda_{j2}$, for $j = 1, 2, \cdots, k$, is not known. We shall, however, consider a choice in analogy with our choice, discussed in Section 1, for the univariate case. Let us choose $\lambda_{j1}, \lambda_{j2}$, for $j = 1, 2, \cdots, k$; so that all the individual regions,

$$\lambda_{j1} \leqq \frac{c_{\min}(S_j)}{c_{\max}(S_0)} \leqq \frac{c_{\max}(S_j)}{c_{\min}(S_0)} \leqq \lambda_{j2}, \tag{4.4}$$

are of the same size $(1 - \alpha^*)$, where α^* is such that the region of intersection, (4.3), is of size $(1 - \alpha)$. Here again, in general, $(1 - \alpha) \neq (1 - \alpha^*)^k$, but we shall assume non-triviality, i.e., given α we can find α^* and vice-versa. As a further condition to determine the λ's completely, let us impose the condition that the individual tests with acceptance regions (4.4) are to be locally unbiased.

Investigations, similar to those of Section 2, for desirable power properties,

which might follow from the second condition on the λ's, have not been made in this inquiry due to the difficulties involved.

A method of evaluating the individual probability integrals like (4.4) is given by the author in [1]. The probability integral associated with (4.3), however, needs to be studied with a view to tabulation.

5. The associated simultaneous confidence bounds on $c(\Sigma_i\Sigma_0^{-1})$ for $i = 1, 2, \cdots, k$. Since $S_0, S_1, \cdots, S_k$ are independently distributed, the joint distribution of the S's is obtained by taking the product of the distributions in (4.2).

Next let us make the following transformations,

$$\Sigma_i(p \times p) = \Lambda_i'(p \times p)\, D_{\gamma_i}(p \times p)\, \Lambda_i(p \times p), \qquad i = 0, 1, 2, \cdots, k \tag{5.1}$$

where each Λ_i is an orthogonal matrix, and the p diagonal elements of each D_{γ_i} are the p characteristic roots, $\gamma_{i1}, \cdots, \gamma_{ip}$, of the corresponding Σ_i (for $i = 0, 1, 2, \cdots, k$).

Then the joint distribution of $S_1, \cdots, S_k$ and S_0 may be rewritten as

$$\text{const.} \prod_{i=0}^{k} \left(\prod_{j=1}^{p} \gamma_{ij}^{\frac{-n_i}{2}} \right) \exp\left[-\tfrac{1}{2} \operatorname{tr}\left\{ \sum_{i=0}^{k} n_i \Lambda_i' D_{1/\gamma_i} \Lambda_i S_i \right\} \right] \prod_{i=0}^{k} |S_i|^{\frac{n_i-p-1}{2}} dS_i,$$

or, remembering that $\operatorname{tr}[A(p \times q)B(q \times p)] = \operatorname{tr}[B(q \times p)A(p \times q)]$, (e.g. [5], p. A-1), as

$$\text{const.} \exp\left[-\tfrac{1}{2} \operatorname{tr}\left\{ \sum_{i=0}^{k} n_i D_{1/\sqrt{\gamma_i}} \Lambda_i S_i \Lambda_i' D_{1/\sqrt{\gamma_i}} \right\} \right] \prod_{i=0}^{k} |S_i|^{\frac{n_i-p-1}{2}} dS_i. \tag{5.2}$$

Let us next make the transformations

$$D_{1/\sqrt{\gamma_i}} \Lambda_i S_i \Lambda_i' D_{1/\sqrt{\gamma_i}} = S_i^*, \qquad i = 0, 1, \cdots, k. \tag{5.3}$$

Then the joint distribution of $S_1^*, \cdots, S_k^*$ and S_0^* is seen to be

$$\text{const.} \exp\left[-\tfrac{1}{2} \operatorname{tr}\left\{ \sum_{i=0}^{k} n_i S_i^* \right\} \right] \prod_{i=0}^{k} |S_i^*|^{\frac{n_i-p-1}{2}} dS_i^*, \tag{5.4}$$

which is of the same form as the joint distribution of $S_1, \cdots, S_k$ and S_0 under H_0.

Therefore, it follows that we can find constants $\lambda_{j1}, \lambda_{j2}$ $(j = 1, 2, \cdots, k)$ such that the simultaneous statements

$$\lambda_{j1} \leqq \frac{c_{\min}(S_j^*)}{c_{\max}(S_0^*)} \leqq \frac{c_{\max}(S_j^*)}{c_{\min}(S_0^*)} \leqq \lambda_{j2}, \qquad \text{for } j = 1, 2, \cdots, k, \tag{5.5}$$

have a joint probability $= (1 - \alpha)$, for a preassigned α. It is well known that all non-zero $c(AB) =$ the non-zero $c(BA)$ (e.g. [5], p. 138). Hence, $c(\Lambda_i S_i \Lambda_i') = c(S_i)$, since Λ_i is orthogonal. Furthermore,

$$c(S_i^*) = c(D_{1/\gamma_i} \Lambda_i S_i \Lambda_i'), \quad \text{where} \quad S_i^*(p \times p),$$

for $i = 0, 1, \cdots, k$, are symmetric and almost everywhere positive definite matrices.

Consider, for any $j = 1, 2, \cdots, k$,

$$\lambda_{j1} \leqq \frac{c_{\min}(S_j^*)}{c_{\max}(S_0^*)} \leqq \frac{c_{\max}(S_j^*)}{c_{\min}(S_0^*)} \leqq \lambda_{j2}, \tag{5.6}$$

which is equivalent to

$$\frac{1}{\lambda_{j1}} \geqq \frac{c_{\max}(S_j^{*-1})}{c_{\min}(S_0^{*-1})} \geqq \frac{c_{\min}(S_j^{*-1})}{c_{\max}(S_0^{*-1})} \geqq \frac{1}{\lambda_{j2}},$$

or to

$$\frac{1}{\lambda_{j1}} \geqq \frac{c_{\max}(D_{\gamma_j} A_j S_j^{-1} A_j')}{c_{\min}(D_{\gamma_0} A_0 S_0^{-1} A_0')} \geqq \frac{c_{\min}(D_{\gamma_j} A_j S_j^{-1} A_j')}{c_{\max}(D_{\gamma_0} A_0 S_0^{-1} A_0')} \geqq \frac{1}{\lambda_{j2}}. \tag{5.7}$$

It is known that if $A_1(p \times p)$ and $A_2(p \times p)$ are two matrices such that A_2 is symmetric positive definite and A_1 is symmetric at least positive semi-definite then $c_{\max}(A_1)\ c_{\max}(A_2) \geqq c(A_1 A_2) \geqq c_{\min}(A_1)\ c_{\min}(A_2)$. Using this, we have $c_{\max}(D_{\gamma_j} A_j S_j^{-1} A_j')\ c_{\max}(A_j S_j A_j') \geqq c_{\max}(D_{\gamma_j})$, so that,

$$c_{\max}(D_{\gamma_j} A_j S_j^{-1} A_j') \geqq c_{\max}(D_{\gamma_j})/c_{\max}(A_j S_j A_j') = c_{\max}(D_{\gamma_j})/c_{\max}(S_j),$$

and similarly,

$$c_{\min}(D_{\gamma_0} A_0 S_0^{-1} A_0') \leqq \frac{c_{\min}(D_{\gamma_0})}{c_{\min}(A_0 S_0 A_0')} = \frac{c_{\min}(D_{\gamma_0})}{c_{\min}(S_0)},$$

and hence we see that the first part of (5.7) implies that

$$\frac{1}{\lambda_{j1}} \geqq \frac{c_{\max}(D_{\gamma_j})}{c_{\min}(D_{\gamma_0})} \frac{c_{\min}(S_0)}{c_{\max}(S_j)}.$$

Again, using the above mentioned result, we note that

$$c_{\min}(D_{\gamma_j} A_j S_j^{-1} A_j') \leqq \frac{c_{\min}(D_{\gamma_j})}{c_{\min}(S_j)}$$

and

$$c_{\max}(D_{\gamma_0} A_0 S_0^{-1} A_0') \geqq \frac{c_{\max}(D_{\gamma_0})}{c_{\max}(S_0)},$$

so that the second part of (5.7) implies that,

$$\frac{c_{\min}(D_{\gamma_j})}{c_{\max}(D_{\gamma_0})} \frac{c_{\max}(S_0)}{c_{\min}(S_j)} \geqq \frac{1}{\lambda_{j2}}.$$

Therefore, we observe that (5.7) implies the statement,

$$\frac{1}{\lambda_{j1}} \frac{c_{\max}(S_j)}{c_{\min}(S_0)} \geqq \frac{c_{\max}(D_{\gamma_j})}{c_{\min}(D_{\gamma_0})} \geqq \frac{c_{\min}(D_{\gamma_j})}{c_{\max}(D_{\gamma_0})} \geqq \frac{1}{\lambda_{j2}} \frac{c_{\min}(S_j)}{c_{\max}(S_0)}. \tag{5.8}$$

We also have

$$\frac{c_{\max}(D_{\gamma_j})}{c_{\min}(D_{\gamma_0})} = c_{\max}(D_{\gamma_j}) c_{\max}(D_{1/\gamma_0}) = c_{\max}(\Sigma_j) c_{\max}(\Sigma_0^{-1}) \geqq c_{\max}(\Sigma_j \Sigma_0^{-1}),$$

and,

$$\frac{c_{\min}(D_{\gamma_j})}{c_{\max}(D_{\gamma_j})} = c_{\min}(D_{\gamma_j})c_{\min}(D_{1/\gamma_0}) = c_{\min}(\Sigma_j)c_{\min}(\Sigma_0^{-1}) \leqq c_{\min}(\Sigma_j\Sigma_0^{-1}).$$

Therefore, we observe that (5.8) implies the confidence statement,

$$\frac{1}{\lambda_{j1}}\frac{c_{\max}(S_j)}{c_{\min}(S_0)} \geqq c_{\max}(\Sigma_j\Sigma_0^{-1}) \geqq c_{\min}(\Sigma_j\Sigma_0^{-1}) \geqq \frac{1}{\lambda_{j2}}\frac{c_{\min}(S_j)}{c_{\max}(S_0)}. \tag{5.9}$$

Combining all statements like (5.9) for $j = 1, 2, \cdots, k$, we see that the statements (5.5) imply the simultaneous confidence statements,

$$\begin{aligned}
&\frac{1}{\lambda_{11}}\frac{c_{\max}(S_1)}{c_{\min}(S_0)} \geqq \text{all } c(\Sigma_1\Sigma_0^{-1}) \geqq \frac{1}{\lambda_{12}}\frac{c_{\min}(S_1)}{c_{\max}(S_0)} \\
&\qquad\vdots \qquad\qquad\qquad\vdots \qquad\qquad\qquad\vdots \\
&\frac{1}{\lambda_{k1}}\frac{c_{\max}(S_k)}{c_{\min}(S_0)} \geqq \text{all } c(\Sigma_k\Sigma_0^{-1}) \geqq \frac{1}{\lambda_{k2}}\frac{c_{\min}(S_k)}{c_{\max}(S_0)}
\end{aligned} \tag{5.10}$$

with a joint confidence coefficient $\geqq (1 - \alpha)$.

6. Concluding remarks. Hartley's test is more involved but has a more detailed structure of alternatives. A generalization of Hartley's test to the case of unequal sample sizes, and a multivariate extension of that, are under investigation and will be discussed in a later paper.

7. Acknowledgement. The author wishes to express his indebtedness to Prof. S. N. Roy for his help and guidance in preparing this paper.

REFERENCES

[1] R. GNANADESIKAN, "Further contributions to multivariate analysis including univariate and multivariate variance components analysis and factor analysis," Institute of Statistics, University of North Carolina, Mimeo. Series No. 158, 1956.

[2] H. O. HARTLEY, "The maximum F-ratio as a short cut test for heterogeneity of variance," *Biometrika*, Vol. 37 (1950), pp. 308–312.

[3] K. V. RAMACHANDRAN, "On the simultaneous analysis of variance test," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 521–528.

[4] S. N. ROY, "On a heuristic method of test construction and its uses in multivariate analysis," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 220–238.

[5] S. N. ROY, *Some Aspects of Multivariate Analysis*, John Wiley and Sons, New York, 1957.

[6] S. N. ROY, "Some further results in simultaneous confidence interval estimation," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 752–761.

[7] S. N. ROY, "A note on 'Some further results in simultaneous confidence interval estimation,'" *Ann. Math. Stat.*, Vol. 27 (1956), pp. 856–858.

[8] S. N. ROY AND R. GNANADESIKAN, "Further contributions to multivariate confidence bounds," *Biometrika*, Vol. 44 (1957), pp. 399–410.

ON THE THEORY OF BAN ESTIMATES[1]

By Robert A. Wijsman[2]

University of California, Berkeley

1. Introduction and summary. The notion of best asymptotically normal estimates—BAN estimates for short[3]—was introduced by Neyman [8] in the multinomial case. Applications have been made in biological problems, notably in bio-assay [2], [4], [5]. Generalizations of Neyman's work have been made by Barankin and Gurland [1], Chiang [3] and Ferguson [5]. The usual theory of BAN estimates requires differentiability of the estimates, and imposes rather strong conditions on certain functions given in advance (the functions ζ and Σ of Section 3). In this note a different definition of BAN estimates is made which does not require differentiability, at the same time relaxing the conditions on ζ and Σ, whereas in essence all important theorems in the theory of BAN estimates are retained.

2. Notation. Convergence in probability is denoted by $\xrightarrow{P}$. $X_n \sim Y_n$ means $X_n - Y_n \xrightarrow{P} 0$. $X_n \xrightarrow{\mathfrak{L}} \mathfrak{N}(0, \Sigma)$ means that the law of X_n tends to a multivariate normal law with mean 0 and covariance matrix Σ. A' is the transpose of a matrix A, I_m the $m \times m$ identity matrix. We shall write *regular* (1), *regular* (2), BAN (1), BAN (2), depending on whether Definition 1 or Definition 2 of Section 3 is used. For notation and terminology not explained here see Chiang's paper [3], with which the notation in this note is in fair agreement.

3. The usual definition and a new definition of BAN estimates. Let Z_n be a sequence of random vectors, taking values in a space $\mathcal{Z}$ which is a subspace of a k-dimensional Euclidean space R^k. The distribution of the Z_n depends on a parameter θ which takes values in an open subset Ω of an m-dimensional Euclidean space, where $m \leqq k$. The true value of θ will be denoted by θ_0. It is assumed that

$$\sqrt{n}\,(Z_n - \zeta(\theta_0)) \xrightarrow{\mathfrak{L}} \mathfrak{N}(0, \Sigma(\theta_0)) \tag{1}$$

in which ζ and Σ are functions on Ω, ζ into $\mathcal{Z}$ and Σ into the space of $k \times k$ positive semi-definite matrices. Let the set $\zeta(\Omega)$ be denoted by U. In the simplest theory of BAN estimates, an admissible estimate is a function $\hat{\theta}$ from $\mathcal{Z}$ to Ω, and if Z_n is observed then θ is estimated by $\hat{\theta}(Z_n)$. We shall occasionally write

Received November 11, 1957; revised November 7, 1958.

[1] This investigation was supported (in part) by a research grant (No. G-3666) from the National Institutes of Health, Public Health Service.

[2] Now at the University of Illinois.

[3] Some authors prefer "RBAN," where "R" stands for *regular*. In this note, regularity of a BAN estimate is part of the definition.

$\hat{\theta}_n$ instead of $\hat{\theta}(Z_n)$. In the usual theory of BAN estimates the following assumptions and definitions are made (our Definition 1 differs slightly from the one given in [3] but leads to the same definition of BAN (1)):

ASSUMPTION 1. (i) ζ is 1-1 and bicontinuous; (ii) $\Sigma(\theta)$ is nonsingular for every θ; (iii) ζ and Σ have continuous second derivatives; (iv) the matrix $\partial\zeta/\partial\theta$ is of rank m for every θ.

DEFINITION 1. $\hat{\theta}$ is called regular (1) if (i) $\hat{\theta}(Z_n) \xrightarrow{P} \theta_0$ whatever θ_0, that is, $\hat{\theta}$ is consistent; (ii) $\partial\hat{\theta}/\partial z$ exists and is continuous in a neighborhood of U.

We shall denote $\zeta(\theta_0) = \zeta_0$, $\Sigma(\theta_0) = \Sigma_0$. The $k \times m$ matrix derivative $\partial\zeta/\partial\theta$ will be denoted by $V(\theta)$, the $m \times k$ matrix derivative $\partial\hat{\theta}/\partial z$ by $A(z)$. $V(\theta_0)$ and $A(\zeta_0)$ will be written V_0, A_0 respectively. An immediate consequence of Assumption 1, Definition 1 and (1) is that

$$\hat{\theta}(\zeta(\theta)) \equiv \theta \text{ identically in } \theta. \tag{2}$$

This enables us to show that $\hat{\theta}_n$ is asymptotically normal, with asymptotic covariance matrix $A_0 \Sigma_0 A_0'$. From (2) one obtains by differentiation

$$A(\zeta(\theta))V(\theta) \equiv I_m \text{ identically in } \theta, \tag{3}$$

which is a very important relation since it has as a consequence that among all matrices of the form $A_0\Sigma_0A_0'$ there is a minimal one. A regular (1) estimate with minimal covariance matrix is called BAN (1).

A priori there is no reason why an estimate should have a derivative, apart from the convenience with which one can show the existence of a minimum covariance matrix among the covariance matrices of regular (1) estimates. The differentiability conditions in Assumption 1(iii) also seem stronger than they need be. The reason for assuming so much differentiability is to ensure the differentiability of $\hat{\theta}$ if the latter is generated by minimizing a quadratic form [3] or by the root of a linear form [5]. Chiang [3] in his theorem 5 makes the full Assumption 1(iii). In his theorem 6 he only assumes that ζ has continuous second derivatives, and makes no assumptions on Σ. However, the type of estimates allowed there goes out beyond the simplest theory of BAN estimates as treated in the present paper, since $\hat{\theta}$ is allowed to be a function of both Z_n and S_n, where S_n is a consistent estimate of Σ. If S_n is taken as a function of Z_n, which is usually the case in applications, it has to be a differentiable function in order that $\hat{\theta}$ is differentiable. Then, since $Z_n \xrightarrow{P} \zeta_0$, we must have $S_n(\zeta_0) = \Sigma_0$, whatever be θ_0; that is, $S_n(\zeta(\theta)) = \Sigma(\theta)$, so that Σ has continuous first derivatives. If a regular (1) estimate is to be obtained as a root of a linear form (equation (5) in the present paper) then the matrix B in this linear form has to have continuous first partial derivatives with respect to z and θ. If, in addition, this regular (1) estimate is to have minimum covariance matrix, then the matrix B has to satisfy $B(\zeta(\theta), \theta) = V'(\theta)\Sigma^{-1}(\theta)$ for all θ (the more general condition given in [5] reduces to the one given here if Σ is non-singular). It follows then that $V'(\theta)\Sigma^{-1}(\theta)$ has to have continuous first derivatives with respect to θ, and

the most natural way to achieve this is to have both V and Σ continuously differentiable. Thus, we see that in order to generate BAN (1) estimates it is practically necessary to require that Σ is continuously differentiable and ζ continuously twice differentiable.

It should be said at once that Ferguson in [5] does not require the estimates to be regular (1), which allows at the same time for relaxation of the conditions on Σ and ζ. The class of estimates considered in [5] consists of all those estimates which can be obtained as a suitably chosen root of a linear form (our equation (5)), for the various choices of the matrix B. The restrictions on B—continuity in z and differentiability in θ—ensure the continuity of $\hat{\theta}$ in a neighborhood of U. The corresponding class of covariance matrices contains the minimum member $(V_0'\Sigma_0^{-1}V_0)^{-1}$, which can be attained if Σ and V are continuous in θ (our Assumption 2(iii)). This minimal covariance matrix has the same form as in the case of regular (1) estimates, so that Ferguson's BAN estimates have the same asymptotic properties as BAN (1) estimates whenever they exist.

It may be argued that from a theoretical point of view it is slightly unsatisfactory to define a class of estimates by the manner in which they are generated, rather than by their common properties. The approach in this paper is different from Ferguson's in that a class of estimates will be defined by the properties they are to have, in the same way as this is done in the case of regular (1) estimates, but with a weaker definition of regular. The new definition of regular (Definition 2, below) will henceforth be referred to as *regular* (2). In spite of this weakening, all important theorems remain valid, in particular the existence of the minimum covariance matrix $(V_0'\Sigma_0^{-1}V_0)^{-1}$. The Assumptions 1(iii) on Σ and ζ will also be relaxed. The only differentiability condition retained is that $V = \partial\zeta/\partial\theta$ exists and is continuous. This condition is certainly a very natural one, since the matrix V plays an important role in the theory of BAN estimates. Thus, compared to the theory of BAN (1) estimates, the field of applicability is enlarged, and at the same time the theory becomes somewhat neater since the conditions are better tuned to the essentials in the theory. The class of regular (2) estimates contains both the regular (1) and Ferguson's estimates. Furthermore, the asymptotic properties of BAN (2) estimates are the same as those of BAN (1) estimates whenever the latter exist. Thus, the class of estimates to be considered is enlarged, and if a BAN (1) estimate exists, it still belongs to the best estimates in this class.

ASSUMPTION 2. (i), (ii) and (iv) are the same as assumptions 1(i), 1(ii) and 1(iv); (iii) Σ is continuous and ζ has a continuous first derivative V.

DEFINITION 2. $\hat{\theta}$ will be called regular (2) if (i) $\hat{\theta}$ is continuous in each point of U; (ii) for each θ there exists an $m \times k$ matrix $A(\theta)$, continuous in θ, such that

$$\sqrt{n}(\hat{\theta}(Z_n) - \theta_0) \sim A(\theta_0)\sqrt{n}(Z_n - \zeta_0) \text{ whatever } \theta_0 . \tag{4}$$

Definition 2 is weaker than Definition 1, in that the former is implied by the latter, but not vice versa. In fact, Definition 2 does not even guarantee the

continuity of $\hat{\theta}$ in a neighborhood of U. This causes a little trouble, since $\hat{\theta}$ is then not necessarily measurable. For the reinterpretation of various probability statements the reader is referred to a remark by LeCam ([6], p. 132). With Definition 2 it is easy to show the validity of (2). The following theorem, however, is not immediate, and will be proved in section 4.

THEOREM 1. *$A(\theta)V(\theta) = I_m$ identically in θ.*

Theorem 1 takes the place of (3), and $A(\theta_0)$ takes the place of what was previously called A_0. Thus, it is seen that the classes of covariance matrices of regular (1) and regular (2) estimates coincide. A regular (2) estimate with minimal covariance matrix will be called BAN (2), and, with a few modifications, the whole theory of BAN estimates is unchanged.

The next theorem is essentially due to Ferguson [5], except for the weaker conditions, and shows how to generate a BAN (2) estimate as a root of a linear form.

THEOREM 2. *Let $B(z, \theta)$ be an $m \times k$ matrix, continuous in θ for each z and continuous in (z, θ) at each point $(\zeta(\theta),\theta)$, such that B_0V_0 is nonsingular whatever θ_0, where $B_0 = B(\zeta_0, \theta_0)$. Then there exists a neighborhood N of U and a function $\hat{\theta}$ on N to Ω such that on N, $\hat{\theta}(z)$ satisfies the equation*

$$B(z, \theta)(z - \zeta(\theta)) = 0 \tag{5}$$

for θ and such that $\hat{\theta}$ is a regular (2) estimate. Furthermore, we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \sim (B_0V_0)^{-1}B_0\sqrt{n}(Z_n - \zeta_0). \tag{6}$$

Lastly, if $B_0 = V_0'\Sigma_0^{-1}$ then $\hat{\theta}$ is BAN.

The proof of Theorem 2 proceeds along conventional lines and will not be reproduced here. It relies on Brouwer's fixed point theorem [7], in which the transformation is assumed to be continuous but not necessarily differentiable, and, as a consequence, uniqueness of the fixed point cannot be concluded. As a result, the estimate $\hat{\theta}$ is not necessarily unique. The matrix B may be chosen to be $V'(\theta)\Sigma^{-1}(\theta)$ and will then satisfy the conditions of Theorem 2, by assumption 2(iii). Moreover, this choice for B will generate a BAN (2) estimate, by the last part of Theorem 2. The same conclusions hold if $V(\theta)$ is replaced by a matrix $V^*(z)$, depending only on z, such that V^* is continuous in each point of U, and $V^*(\zeta(\theta)) = V(\theta)$ for all θ. $\Sigma(\theta)$ may be replaced similarly.

4. Proof of Theorem 1. We shall need the following lemma, whose proof is due to Dr. Lucien M. LeCam.

LEMMA 1. *Let (a, b) be a one-dimensional interval in R^k, and suppose that for each $x \in (a, b)$ there is a sphere $S(x)$ about x with radius $r(x)$. Then there are two distinct points, x_1 and x_2, in (a, b) such that $x_1 \in S(x_2)$ and $x_2 \in S(x_1)$.*

PROOF. Let $x_0 \in (a, b)$ be a point such that there is no $x \in (a, b) \cap S(x_0)$ for which $x_0 \in S(x)$. Then $r(x) \to 0$ as $x \to x_0$. On the other hand, $r(x_0) > 0$. It follows that r has a discontinuity of the first kind at x_0. The conclusion follows from the fact that the discontinuities of the first kind are denumerable.

We proceed to prove Theorem 1. Suppose that for some $\theta_1 \in \Omega$, $A(\theta_1)V(\theta_1) \neq I_m$.

Without loss of generality we may assume that $\zeta(\theta_1)$ is the origin 0 of the coordinate system in Z. Let T be the tangent plane to U in 0, and consider in the following T fixed, i.e. not subject to transformations. There is a transformation $z \to z^* = g(z)$ which in a Z-neighborhood of 0 is continuously differentiable in both directions, with $\partial g(0)/\partial z = I_k$, and which maps U into T. More precisely, there is a U-neighborhood N_u of 0, a T-neighborhood N_t of 0, and a Z-neighborhood N_z of N_u, such that on N_z the transformation g has the differentiability properties mentioned in the preceding sentence, and such that $g(N_u) = N_t$. Let $N_\theta = \zeta^{-1}(N_u)$. In (4) we shall only consider values of θ_0 which are in N_θ. The estimate $\hat{\theta}$ is now to be considered as a function of z^*. If we write (4) down for the transformed variables, i.e. replacing Z_n by Z_n^*, ζ_0 by ζ_0^*, we have to replace $A(\theta_0)$ by $A^*(\theta_0) = (\partial g(\zeta_0)/\partial z)^{-1} A(\theta_0)$. $A^*(\theta)$, like $A(\theta)$, is continuous in θ. Since $\partial g(\zeta(\theta_1))/\partial z = I_k$, we have $A^*(\theta_1) = A(\theta_1)$. Furthermore, for the transformed variables we have $V^*(\theta) = \partial\zeta^*/\partial\theta$, where $\zeta^*(\theta) = g(\zeta(\theta))$, so that $V^* = (\partial g/\partial z)V$ (the arguments have been suppressed). At θ_1 we have $V^*(\theta_1) = V(\theta_1)$. Thus, if $A(\theta_1)V(\theta_1) \neq I_m$, then also $A^*(\theta_1)V^*(\theta_1) \neq I_m$. Dropping the asterisks, we consider a new problem, in which the new Ω is the old N_θ, the new U is the old N_t. Hence U is a subset of an m-dimensional subspace of R^k. For some $\theta_1 \varepsilon \Omega$, $A(\theta_1)V(\theta_1) \neq I_m$. We may further simplify the problem by making a suitable transformation $\theta \to \theta^*$, continuously differentiable in both directions. This transforms $V \to V^* = V(\partial\theta/\partial\theta^*)$ and $A \to A^* = (\partial\theta^*/\partial\theta)A$ (the arguments have been suppressed). Thus, if $AV \neq I_m$, then also $A^*V^* \neq I_m$. We choose the transformation $\theta \to \theta^* = \zeta(\theta)$. Dropping the asterisks, in the new problem Ω and U are identical, ζ is the identity function on U, and $\hat{\theta}$ is a function on Z to U which is the identity function on U, by (2). For each $u \varepsilon U$, $A(u)$ is a linear transformation of Z into the m-dimensional subspace in which U is embedded. If Theorem 1 is true, then, for each u, $A(u)$ is the identity transformation on U. We have assumed that $A(0)$ is not the identity transformation on U, and will show that this leads to a contradiction.

If $A(0)$, restricted to U, is not the identity transformation, then the same is true on at least one of the coordinate axes in U. We choose one of these coordinate axes, and call it the x-axis in the following. For simplicity, x, with or without subscripts, will stand both for a point on the x-axis and for its x-coordinate. In the following, $|z|$ will denote the norm of a vector z, $\|A\|$ the norm of a matrix A, I the identity transformation on U. Using the continuity of A, there is on the x-axis an interval $N_a = (-a, a)$ and there is an $\epsilon > 0$ such that for all $\xi, x_1, x_2 \varepsilon N_a$ we have

$$|(A(\xi) - I)(x_2 - x_1)| \geqq 6\epsilon\,|x_2 - x_1|\,. \tag{7}$$

Furthermore, we can choose a so small that for all $x_1, x_2 \varepsilon N_a$ we have

$$\|A(x_2) - A(x_1)\| < \epsilon. \tag{8}$$

We now put

$$f(z, u) = \hat{\theta}(z) - u - A(u)(z - u), \tag{9}$$

then from (4) it follows that

$$P\{\sqrt{n}\,|f(Z_n, u_0)| \geqq \epsilon\} \to 0, \tag{10}$$

in which u_0 is the true value of the parameter. Denote by $S(u, n)$ the open sphere with radius $n^{-1/2}$ about u. If $u = 0$ we write simply $S(n)$. Using (10) we have

$$P\{Z_n \in S(u_0, n), \sqrt{n}\,|f(Z_n, u_0)| \geqq \epsilon\} \to 0. \tag{11}$$

Making the transformation $y = n^{1/2}(z - u_0)$, $Y_n = n^{1/2}(Z_n - u_0)$, $g_n(y) = n^{1/2}f(z, u_0)$, we can write (11) as

$$P\{Y_n \in S(1), |g_n(Y_n)| \geqq \epsilon\} \to 0. \tag{12}$$

Let μ be the Lebesgue measure in R^k. Since Y_n has a limiting density with respect to μ, we conclude from (12)

$$\mu\{y \in S(1), |g_n(y)| \geqq \epsilon\} \to 0. \tag{13}$$

If we divide the left-hand side of (13) by the constant $\mu S(1) = \mu\{y \in S(1)\}$, and make the transformation back from y to z, we obtain

$$[\mu S(n)]^{-1}\mu\{z \in S(u, n), |f(z, u)| \geqq \epsilon n^{-1/2}\} \to 0 \tag{14}$$

in which we have dropped the subscript 0 on u.

For any z, let $z = x_z + y_z$, where x_z is the x-component of z. We choose a number $\alpha > 0$ in such a way that for any x_1, x_2, n, with

$$\tfrac{1}{2}n^{-1/2} < x_2 - x_1 < n^{-1/2} \tag{15}$$

we have

$$\mu\{z \in S(x_1, n) \cap S(x_2, n), x_1 < x_z < x_2\} > \alpha\mu S(n). \tag{16}$$

The number α can obviously be chosen independently of the particular choice of x_1, x_2, n, so long as (15) holds. Restricting now u in (14) to points x in N_ϵ, we have for each $x \in N_\epsilon$ an integer n_x such that

$$\mu\{z \in S(x, n), |f(z, x)| \geqq \epsilon n^{-1/2}\} < \tfrac{1}{2}\alpha\mu S(n) \tag{17}$$

provided $n \geqq n_x$. According to Lemma 1, there are two points $x_1, x_2 \in N_\epsilon$, with $x_1 < x_2$, such that $x_1 \in S(x_2, n_{x_2})$ and $x_2 \in S(x_1, n_{x_1})$. We can now choose an integer $n \geqq \max(n_{x_1}, n_{x_2})$ such that (15) holds, and therefore also (16). The two equations (16) and (17) together imply that

$$\mu\{z \in S(x_1, n) \cap S(x_2, n), x_1 < x_z < x_2, \\ |f(z, x_1)| < \epsilon n^{-1/2}, |f(z, x_2)| < \epsilon n^{-1/2}\} > 0, \tag{18}$$

so that the set in braces in (18) is not empty. Choosing any point z in this set, and using (15), we have the following inequalities:

$$x_1 < x_z < x_2, \tag{19}$$

$$|y_s| < 2(x_2 - x_1), \tag{20}$$

$$|f(z, x_1)| < 2\epsilon(x_2 - x_1),\ |f(z, x_2)| < 2\epsilon(x_2 - x_1). \tag{21}$$

From (9) we compute

$$\begin{aligned} f(z, x_2) - f(z, x_1) &= (A(x_1) - I)(x_s - x_1) \\ &\quad + (A(x_2) - I)(x_2 - x_s) + (A(x_1) - A(x_2))y_s . \end{aligned} \tag{22}$$

By continuity of A and by virtue of (19) there is a point $\xi \in (x_1, x_2)$ such that

$$|(A(\xi) - I)(x_2 - x_1)| = |(A(x_1) - I)(x_s - x_1) + (A(x_2) - I)(x_2 - x_s)|. \tag{23}$$

Using (8), (20), (21), (22) and (23), we have finally

$$|(A(\xi) - I)(x_2 - x_1)| < 6\epsilon(x_2 - x_1), \tag{24}$$

contradicting (7). Q.E.D.

Acknowledgment. The author is indebted to Dr. Lucien M. LeCam for suggesting the proof of Lemma 1, and to the referee for helpful comments.

REFERENCES

[1] E. W. Barankin and J. Gurland, "On asymptotically normal, efficient estimators I," *University of California Publications in Statistics*, Vol. 1 (1950), pp. 89–130.

[2] J. Berkson, "Application of the logistic function to bio-assay," *J. Am. Stat. Assn.*, Vol. 39 (1944), pp. 357–365.

[3] C. L. Chiang, "On regular best asymptotically normal estimates," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 336–351.

[4] C. L. Chiang, "An application of stochastic processes to experimental studies on flour beetles," *Biometrics*, Vol. 13 (1957), pp. 79–97.

[5] T. S. Ferguson, "A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities," Dissertation, University of California, Berkeley, 1956, *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1046–1062.

[6] L. M. LeCam, "On the asymptotic theory of estimation and testing hypotheses," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1956, pp. 129–156.

[7] S. Lefschetz, *Introduction to Topology*, Princeton University Press, 1949.

[8] J. Neyman, "Contribution to the theory of the χ^2 test," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1949, pp. 239–273.

ESTIMATION OF THE MEDIANS FOR DEPENDENT VARIABLES

By Olive Jean Dunn[1]

Statistical Laboratory, Iowa State College

1. Summary. Joint intervals of bounded confidence are suggested for the medians of a bivariate population with continuous marginal distributions. The two intervals are of the classic type based on sample order statistics.

2. Introduction. The problem considered in this paper is that of using a nonparametric method to estimate by a confidence set the unknown medians of two dependent variables. In various types of research, it is convenient to consider a sample of n individuals and to take measurements on the same n individuals at two different times or at two different levels of treatment. The two measurements on the same individual cannot be assumed to be independent, so that it is appropriate to consider the $2n$ measurements as a sample of size n from a bivariate distribution.

Let the two variables y_1, y_2 with medians ν_1, ν_2 have the c.d.f. (cumulative distribution function) $F(y_1, y_2)$. By a set of simultaneous confidence intervals of bounded confidence level $1 - \alpha$ for ν_1, ν_2 is meant a set of four functions of the sample values, say g_1, g_2, h_1, h_2, such that

$$P(g_1 < \nu_1 < h_1, g_2 < \nu_2 < h_2) \geqq 1 - \alpha.$$

The probability relationship must hold for all underlying distributions in a specified set of distributions. In this paper the specified set will consist of all bivariate distributions whose marginals have continuous c.d.f.'s.

The method used in this paper to obtain confidence intervals uses order statistics and requires only the assumption that the marginal distributions be continuous.

3. Confidence intervals for the medians of a bivariate distribution. Let the c.d.f. of the variables y_1, y_2 be $F(y_1, y_2)$ and let the two marginal distributions be $F_1(y_1)$ and $F_2(y_2)$, both of which are continuous.

A random sample of n observations will be denoted by $(y_{11}, y_{21}), \cdots, (y_{1n}, y_{2n})$. For $i = 1$ and 2, the set $y_{i1}, \cdots, y_{in}$ will be reordered from smallest to largest and renamed $z_{i1}, \cdots, z_{in}$. Thus $z_{i1} \leqq z_{i2} \leqq \cdots \leqq z_{in}$ for $i = 1$ or 2. The z_{1j} and z_{2j} need not belong to the same observation.

Two positive integers, r and s, such that $2r + s = n$, are selected. Let E_i be the event that $z_{ir} < \nu_i < z_{i,r+s+1}$ for $i = 1, 2$. Then, for $i = 1, 2$,

$$P(E_i) = \sum_{j=r}^{r+s} \binom{n}{j} (\tfrac{1}{2})^n = (1 - \alpha)^{1/2}, \quad \text{say.} \tag{1}$$

Received February 25, 1957; revised June 20, 1958.

[1] Now at the University of California at Los Angeles.

If the variables were independent, probabilities could be multiplied, so that $P(E_1E_2) = 1 - \alpha$. This would give a set of intervals of exact confidence level for ν_1 and ν_2, namely z_{1r} to $z_{1,r+s+1}$ for ν_1, z_{2r} to $z_{2,r+s+1}$ for ν_2.

For dependent variables, the same set of intervals may be used as a set with bounded confidence level for it can be shown that $P(E_1E_2) \geqq 1 - \alpha$. It should be noted that symmetric order statistics have been used, and indeed the result would not hold otherwise.

The following proof establishes the necessary inequality.

THEOREM. $P(E_1E_2) \geqq P(E_1)P(E_2)$.

PROOF. Since $P(E_1E_2) = P(E_2 \mid E_1)P(E_1)$, it will be sufficient to prove that $P(E_2 \mid E_1) \geqq P(E_2)$.

If for a certain observation, $y_{1j} > \nu_1$, then let the conditional probability that $y_{2j} > \nu_2$ (this will be referred to as the probability of a "success") be denoted by p. Then using the fact that $F_1(\nu_1) = F_2(\nu_2) = \frac{1}{2}$:

$$p = P(y_{2j} > \nu_2 \mid y_{1j} > \nu_1) = 2F(\nu_1, \nu_2). \tag{2}$$

Similarly, if it is known that $y_{1j} < \nu_1$, then let q be the conditional probability that $y_{2j} > \nu_2$ (probability of a "success"). Then

$$q = P(y_{2j} > \nu_2 \mid y_{1j} < \nu_1) = 1 - 2F(\nu_1, \nu_2). \qquad \text{Thus, } p + q = 1. \tag{3}$$

If it is known that E_1 has occurred, then $r + i$ observations have $y_{1j} < \nu_1$, and so have a conditional probability of success of q, where i may be $0, 1, \cdots s$; $r + s - i$ observations have $y_{1j} > \nu_1$, and so have a conditional probability of success of p.

To obtain a generating function for the probabilities of various numbers of successes, conditioned by the fact that $y_{1r} < \nu_1 < y_{1,r+s+1}$, one may proceed as follows. Let $E_1(i)$ be the event that E_1 occurs with $r + i$ observations having $y_{1j} < \nu_1$. Then $E_1 = \bigcup_0^s E_1(i)$.

Let $Y_<$ be the number of successes in the $r + i$ observations which have $y_{1j} < \nu_1$; let $Y_>$ be the number of successes in the $r + s - i$ observations which have $y_{1j} > \nu_2$; let $Z = Y_< + Y_>$ be the total number of successes. Then

$$\begin{aligned} E\left(t^z \mid \bigcup_0^s E_1(i)\right) &= \sum_{i=0}^{s} E(t^z \mid E_1(i)) \cdot P(E_1(i) \mid E_1) \\ &= \sum_{i=0}^{s} E(t^z \mid E_1(i)) \cdot P(E_1(i))/P(E_1). \end{aligned} \tag{4}$$

Since under the condition $E_1(i)$, $Y_<$ and $Y_>$ are independent,

$$E\left(t^z \mid \bigcup_0^s E_1(i)\right) = \sum_{i=1}^{s} E(t^{Y<} \mid E_1(i))E(t^{Y>} \mid E_1(i))P(E_1(i))/P(E_1). \tag{5}$$

The generating function G for the conditional probabilities of various num-

bers of successes, given that $z_{1r} < \nu_1 < z_{1,r+s+1}$, is, finally,

$$G = C \sum_{i=0}^{s} \frac{1}{(r+i)!(r+s-i)!} (p+qt)^{r+i}(q+pt)^{r+s-i}, \tag{6}$$

where

$$C = 1 \Big/ \sum_{i=0}^{s} \frac{1}{(r+i)!(r+s-i)!}.$$

If a_j, $j = 0, 1, \cdots, n$, is defined as the coefficient of t^j in G, then

$$G = \sum_{j=0}^{n} a_j t^j, \tag{7}$$

and

$$P(E_2 \mid E_1) = \sum_{j=r}^{r+s} a_j. \tag{8}$$

The task now is to show that $P(E_2 \mid E_1)$ has a minimum at $p = \frac{1}{2}$. To do this, the derivative of G is obtained indirectly by differentiating equation (6) with respect to p, and then manipulating the derivative, G_p, as follows:

$$\begin{aligned} G_p &= C \sum_{i=0}^{s} \frac{(1-t)}{(r+i)!(r+s-i)!} [(r+i)(p+qt)^{r+i-1}(q+pt)^{r+s-i} \\ &\qquad - (r+s-i)(p+qt)^{r+i}(q+pt)^{r+s-i-1}] \\ &= \frac{C(1-t)}{(r-1)!(r+s)!} (p+qt)^{r-1}(q+pt)^{r-1}[(q+pt)^{s+1} - (p+qt)^{s+1}]. \end{aligned} \tag{9}$$

Let $C^* = (C/(r-1)!(r+s)!)$. Then

$$\begin{aligned} G_p &= C^*(1-t)(p+qt)^{r-1}(q+pt)^{r-1}[(q^{s+1} - p^{s+1}) + \tbinom{s+1}{1}pq(q^{s-1} - p^{s-1})t \\ &\quad + \tbinom{s+1}{2}p^2q^2(q^{s-3} - p^{s-3})t^2 + \cdots + \tbinom{s+1}{2}p^2q^2(p^{s-3} - q^{s-3})t^{s-1} \\ &\quad + \tbinom{s+1}{1}pq(p^{s-1} - q^{s-1})t^s + (p^{s+1} - q^{s+1})t^{s+1}] \\ &= C^*(1-t)(q-p)(p+qt)^{r-1}(q+pt)^{r-1}[b_0(1-t^{s+1}) \\ &\quad + b_1t(1-t^{s-1}) + b_2t^2(1-t^{s-3}) + \cdots] \\ &= C^*(1-t)^2(q-p)(p+qt)^{r-1}(q+pt)^{r-1}[b_0 + (b_0+b_1)t \\ &\quad + (b_0+b_1+b_2)t^2 + \cdots + (b_0+b_1+b_2)t^{s-2} + (b_0+b_1)t^{s-1} + b_0t^s]. \end{aligned} \tag{10}$$

Here $b_j = \binom{s+1}{j}p^jq^j(q^{s-2j} + q^{s-2j-1}p + \cdots + p^{s-2j})$; hence the partial sums of the b_j's appearing as coefficients within the preceding square brackets are positive and (excluding the trivial cases $p = 0$ or 1) they increase toward the center coefficient(s).

Multiplication of the polynomial within the square brackets by

$$(p+qt)^{r-1}(q+pt)^{r-1}$$

can be accomplished by successive multiplications by $(p + qt)(q + pt)$, and it is easily verified that each multiplication yields a polynomial whose coefficients increase toward the center coefficient(s). Thus

$$(11)\quad G_r = C^*(q - p)(1 - 2t + t^2)$$
$$(c_0 + c_1t + c_2t^2 + \cdots + c_2t^{n-4} + c_1t^{n-3} + c_0t^{n-2})$$
$$= C^*(q - p)(d_0 + d_1t + d_2t^2 + \cdots + d_2t^{n-2} + d_1t^{n-1} + d_0t^n).$$

Here $d_j = c_j - 2c_{j-1} + c_{j-2}$, for $j = 0, 1, \cdots$, and it is understood that $c_{-1} = c_{-2} = 0$.

The derivative with respect to p of the conditional probability $P(E_2 \mid E_1)$ is then $\sum_{j=r}^{r+s} C^*(q - p)\, d_j = 2C^*(q - p)(c_{r-2} - c_{r-1})$. Since $c_{r-2} - c_{r-1}$ is always negative, the derivative is positive, zero, or negative according to whether $p > \frac{1}{2}$, $p = \frac{1}{2}$ or $p < \frac{1}{2}$. Thus $P(E_2 \mid E_1)$ has a minimum at $p = \frac{1}{2}$, and $P(E_1E_2) \geqq P(E_1)P(E_2)$.

Since $E_1E_2 \subset E_1$, one may further write $P(E_1) \geqq P(E_1E_2) \geqq P(E_1)P(E_2)$. The confidence level for the intervals z_{1r} to $z_{1,r+s+1}$, z_{2r} to z_{2r} to $z_{2,r+s+1}$ can actually be as high as $(1 - \alpha)^{1/2}$ (for a distribution function such that $p = 0$ or $p = 1$) or as low as $(1 - \alpha)$ (when $p = \frac{1}{2}$).

4. Evaluation. One way to compare sets of confidence intervals is on the basis of their lengths, or the expected values of their lengths. I shall exhibit some length comparisons when y_1, y_2 are jointly normally distributed with means (or medians) ν_1, ν_2, variances σ_1^2, σ_2^2, and arbitrary co-variance. In Table I, the intervals for the medians obtained by the method of this paper (Method I) are compared with three sets of intervals for the means of a bivariate normal distribution (Methods II and III and IV) obtained in another paper [1]. It should be mentioned that all four methods lead to intervals of bounded confidence.

The figures given in the body of the table are values of $(n^{1/2}/\sigma_i)E(\frac{1}{2}l_i)$, where l_i is the length of the confidence interval for ν_i, $i = 1$ or 2. Values of $1 - \alpha$ (which for Method I cannot be chosen arbitrarily) have been selected as close as possible to .95.

Method II (section 4.2 in [1]), which uses Hotelling's T distribution, is similar to Method I in that no assumptions need be made concerning the variances. When n is small, the intervals are seen to be slightly longer on the average for II than for I.

Method III, based on the Student t distribution, requires that the variances be equal, though they may be unknown. These intervals are seen to be somewhat shorter than those from Method I. Method III is found in section 7.2 of [1].

In Method IV, the variances are assumed to be known and the intervals obtained (the method of section 7.1 in [1]) are the shortest possible intervals for means of the bivariate normal distribution when nothing is known about the

TABLE I

Comparison of Expected Values of Lengths of Confidence Intervals for the Means of a Bivariate Normal Distribution

$\frac{\sqrt{n}}{\sigma_i} E(\frac{1}{2}l_i)$ for:

n	$1-\alpha$	$(1-\alpha)^{1/2}$	Method I: Order Statistics		Method II: Hotelling's T	Method III: Variances Unknown but Equal	Method IV: Variances Known
6	.939	.969	z_{i1} to z_{i5}	3.10	3.72	2.83	2.16
8	.984	.992	z_{i1} to z_{i8}	4.03	4.40	3.54	2.65
10	.958	.979	z_{i2} to z_{i9}	3.17	3.21	2.72	2.31
20	.976	.988	z_{i3} to z_{i18}	3.33	3.12	2.77	2.51
100	.958	.979	z_{i30} to z_{i72}	2.93	2.57	2.33	2.31

l_i = length of the confidence interval for μ_i (or ν_i), $i = 1, 2$.

Method I: $E(\frac{1}{2}l_i)$ computed using the expected value of order statistics in [2].

Method II: The intervals are $\bar{y}_i \pm (\hat{\sigma}_i/\sqrt{n})c_\alpha$, $i = 1, 2$, where c_α is the $1-\alpha$ point in the distribution of Hotelling's T, and $\hat{\sigma}_i^2 = \sum_{j=1}^{n}(y_{ij} - \bar{y}_i)^2/(n-1)$, $i = 1, 2$.

Method III: The intervals are $\bar{y}_i \pm (\hat{\sigma}_1/\sqrt{n})c_\alpha$, $i = 1, 2$, where c_α is the

$$[1 + (1-\alpha)^{1/2}]/2$$

point of the Student t distribution with $n-1$ degrees of freedom, and

$$\hat{\sigma}_1^2 = \sum_{j=1}^{n}(y_{1j} - \bar{y}_1)^2/(n-1)$$

Method IV: The intervals are $\bar{y}_1 \pm (\hat{\sigma}_i/\sqrt{n})\cdot c_\alpha$, $i = 1, 2$, where c_α is the

$$[1 + (1-\alpha)^{1/2}]/2$$

point of the standard normal distribution.

covariance. They thus make a useful standard for purposes of comparison. These "best" intervals are considerably shorter than those from Method I, but this must be balanced against the fact that for the latter method no assumptions are necessary concerning the variance or concerning distributional form (except for continuity).

The fact that Method I, assuming nothing about the form of the distribution, gives shorter intervals for small n than Method II, which demands a normal distribution may seem somewhat surprising. The explanation lies in the fact that, for a given value of α, the actual probability of coverage is higher for Method II than for Method I. For $\rho = 0$ and $\rho = 1$, the actual probabilities of coverage for all four methods are as follows:

n	Methods I, III, IV		Method II	
	$\rho = 0$	$\rho = 1$	$\rho =$	$\rho = 1$
6	.939	.969	.989	.994
8	.984	.992	.997	.999
10	.958	.979	.991	.995
20	.976	.988	.994	.997
100	.958	.979	.988	.994

Throughout the preparation of this paper, it was conjectured that the same set of intervals developed here might be used for a k-variate distribution. Mr. Ernest V. Scheuer has, however, recently drawn the author's attention to the following counterexample, which shows that it is possible for $P(E_1E_2E_3)$ to be less than $P^3(E_1)$, where E_i is the event that $z_{ir} < \nu_i < z_{i,r+s+1}$, $i = 1, 2, 3$.

Let $r = 1$, and, for simplicity, let $\nu_1 = \nu_2 = \nu_3 = 0$.

Let $p_{111} = P(y_1 > 0, y_2 > 0, y_3 > 0)$, $p_{110} = P(y_1 > 0, y_2 > 0, y_3 < 0)$, and similarly for p_{101}, p_{011}, p_{100}, p_{010}, p_{001}, and p_{000}.

It can be readily verified that $P(z_{11} < 0 < z_{1n}, z_{21} < 0 < z_{2n}, z_{31} < 0 < z_{3n})$ is smaller for $p_{111} = p_{100} = p_{010} = p_{001} = 1/4$, $p_{110} = p_{101} = p_{011} = p_{000} = 0$ than it is under independence ($p_{111} = p_{110} = \cdots = p_{000} = 1/8$).

REFERENCES

[1] Olive Jean Dunn, "Estimation of the means of dependent variables," *Ann. Math. Stat.*, Vol. 29 (December 1958), pp. 1095–1111.

[2] D. Teichroew, "Tables of expected values of order statistics and products of order statistics for samples of size twenty or less from the normal distribution," *Ann. Math. Stat.* Vol. 27 (June 1956), pp. 410–426.

A CONSISTENT ESTIMATOR OF A COMPONENT OF A CONVOLUTION

By William R. Gaffey

University of California

1. Introduction and summary. Suppose the observed random variable X is the sum of two independent random variables Z and Y, where Z has a normal distribution with zero expectation and a known variance, and Y has a distribution function, say $G(y)$, which is completely unknown. Then the distribution function of X may be written as

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} G(y) \exp\left[- \frac{(x-y)^2}{2\sigma^2} \right] dy, \tag{1.1}$$

where $F(x)$ and $G(y)$ are unknown.

We consider here the problem of estimating $G(y)$ from a sample $x_1, x_2, \cdots x_n$. Such a problem may arise if, for example, each x_i represents a serum cholesterol determination on one human being randomly selected from some population. Then each x_i may be thought of as the true cholesterol value for that person, plus an "instrumental error" introduced by the complex chemical analysis. We may wish to "correct" for the instrumental error, so to speak, by estimating the distribution of true cholesterol levels in the population.

The maximum likelihood and minimum distance principles do not seem to yield estimators which may be expressed as explicit, more or less easily computable functions of the sample values. We present such an estimator, which is consistent at every continuity point of $G(y)$. (We consider only continuity points throughout the paper.) The estimator is constructed by first exhibiting an inversion formula for $G(y)$ in terms of the derivatives of $F(x)$, and then replacing the derivatives by the difference quotients of the empiric distribution function $F_n(x)$.

The asymptotic mean square error of the estimator is derived, and a rough rule is suggested for deciding when it is worthwhile to compute an estimate. The fact that the estimator is still consistent under certain kinds of dependence between Z and Y is indicated. Finally, some comments are made on the relationship between the present estimator and one derived by Eddington.

2. The inversion formula for $G(x)$. Denote by $F^{(2k)}(x)$ the $2k$th derivative of $F(x)$ evaluated at x. Pollard [3] has derived a formula which states, in effect, that

$$\lim_{n \to 0} \sum_{k=0}^{\infty} (-1)^k F^{(2k)}(x) (\sigma^2 t_n/2)^k / k! = G(x), \tag{2.1}$$

where $\{t_n\}$ is any increasing sequence of positive numbers with $\lim t_n = 1$.

We require a modification of this inversion formula. Let $\{t_n\}$ be an increasing

Received November 26, 1957; revised September 20, 1958.

sequence of positive numbers such that $\lim t_n = 1$ and $\lim n^k t_n^n = 0$ for any $k > 0$. Consider

$$G_n(x) = \sum_{k=0}^{n} (-1)^k F^{(2k)}(x)(\sigma^2 t_n/2)^k/k!. \tag{2.2}$$

Then $\lim G_n(x) = G(x)$. To prove this it is sufficient to show that

$$\lim \sum_{k=n+1}^{\infty} (-1)^k F^{(2k)}(x)(\sigma^2 t_n/2)^k/k! = 0. \tag{2.3}$$

Now

$$F^{(2k)}(x) = \frac{1}{\sqrt{\pi}\,(\sqrt{2}\sigma)^{2k+1}} \int_{-\infty}^{\infty} H_{2k}\left(\frac{x-y}{\sqrt{2}\sigma}\right) \exp\left[-\frac{(x-y)^2}{2\sigma^2}\right] G(y)\,dy, \tag{2.4}$$

where $H_{2k}(x)$ is the $2k$th Hermite polynomial satisfying

$$H_{2k}(x) \leqq A 2^k \sqrt{(2k)!}\, \exp[x^2/2], \tag{2.5}$$

A being independent of x and k. ([5], p. 236)

Therefore,

$$F^{(2k)}(x) \leqq A\sqrt{(2k!)}/\sigma^{2k}, \tag{2.6}$$

and consequently the absolute value of the sum in (2.3) is at most equal to

$$A \sum_{k=n+1}^{\infty} \sqrt{(2k)!}\,(t_n/2)^k/k!. \tag{2.7}$$

Since the coefficients of t_n^k are bounded above, the absolute value of (2.7) is no greater than

$$A t_n^{n+1}(1 - t_n)^{-1}, \tag{2.8}$$

which approaches zero faster than any negative power of n.

If $G^{(r)}(x)$, the rth derivative of $G(x)$, is integrable, and continuous at a particular value of x, a similar argument shows that $\lim G_n^{(r)}(x) = G^{(r)}(x)$.

For later use, we note that by virtue of (2.6) and the continuity of any derivative of $F(x)$,

$$F^{(2k)}(a) = F^{(2k)}(b) + O(a - b)\sqrt{(2k)!}/\sigma^{2k}, \tag{2.9}$$

and that the upper bound of $O(a - b)$ is independent of k.

3. The estimator of $G(x)$ and its mean square error. Define the $2k$th difference quotient of the empiric distribution function $F_n(x)$ by

$$F_n^{(2k)}(x, h) = (2h)^{-2k} \sum_{j=0}^{2k} \binom{2k}{j} (-1)^j F_n(x + (k - j)2h), \tag{3.1}$$

for $h > 0$. The estimator of $G(x)$, for a sample of size n, will then be

$$\hat{G}_n(x) = \sum_{k=0}^{n} (-1)^k F_n^{(2k)}(x, h_n)(\sigma^2 t_n/2)^k/k!, \tag{3.2}$$

where $\lim h_n = 0$.

Clearly, the asymptotic properties of the estimator will depend on the choice of sequences $\{h_n\}$ and $\{t_n\}$. We derive below an asymptotic expression for the mean square error on the assumption that $G^{(4)}(x)$ is integrable, and continuous at the particular value of x involved. Where there is no possibility of confusion, the subscripts are omitted from h_n and t_n.

Consider first the expectation of the $2k$th difference quotient.

$$(3.3)\qquad (2h)^{2k}E[F_n^{(2k)}(x,h)] = \int_{x-h}^{x+h}\cdots\int_{y_{2k-1}-h}^{y_{2k-1}+h} F^{(2k)}(y_{2k})\,dy_{2k}\cdots dy_1.$$

We may write

$$(3.4)\qquad \begin{aligned} F^{(2k)}(y_{2k}) &= F^{(2k)}(y_{2k-1}) + (y_{2k}-y_{2k-1})F^{(2k+1)}(y_{2k-1}) \\ &\quad + (1/2)(y_{2k}-y_{2k-1})^2F^{(2k+2)}(\xi), \end{aligned}$$

where ξ is between y_{2k} and y_{2k-1}. When the integral with respect to y_{2k} is taken, the second term on the right vanishes. Complete integration of the third term on the right results in an expression which may be written as

$$(3.5)\qquad (2h)^{2k}F^{(2k+2)}(\xi)h^2/6,$$

where $x - 2hk \leqq \xi \leqq x + 2hk$. Therefore,

$$(3.6)\qquad \begin{aligned} (2h)^{2k}E[F_n^{(2k)}(x,h)] &= 2h\int_{x-h}^{x+h}\cdots\int_{y_{2k-2}-h}^{y_{2k-2}+h} F^{(2k)}(y_{2k-1})\,dy_{2k-1}\cdots dy_1 \\ &\quad + (2h)^{2k}F^{(2k+2)}(\xi)h^2/6. \end{aligned}$$

Repeating the process, we find finally that

$$(3.7)\qquad (2h)^{2k}E[F_n^{(2\ )}(x,h)] = (2h)^{2k}F^{(2k)}(x) + (2h)^{2k}\sum_{i=1}^{2k} F^{(2k+2)}(\xi_i)h^2/6,$$

where $x - 2hk \leqq \xi_i \leqq x + 2hk$ for all i, or more simply,

$$(3.8)\qquad E[F_n^{(2k)}(x.\ h)] = F^{(2k)}(x) + 2kF^{(2k+2)}(\eta_k)h^2/6,$$

where $x - 2hk \leqq \eta_k \leqq x + 2kh$. Making this substitution in the expectation of $\hat{G}_n(x)$, we have

$$(3.9)\qquad E[\hat{G}_n(x)] = G_n(x) + \frac{h^2}{6}\sum_{k=0}^{n}(-1)^kF^{(2k+2)}(\eta_k)(\sigma^2t/2)^k 2k/k!.$$

Applying (2.9) we obtain, after some algebra,

$$(3.10)\qquad \begin{aligned} E[\hat{G}_n(x)] &= G_n(x) - \frac{h^2\sigma^2t}{6}\sum_{k=0}^{n-1}(-1)^kF^{(2k+4)}(x)(\sigma^2t/2)^k/k! \\ &\quad + O\left[h^3\sum_{k=1}^{n} k\sqrt{(2k+2)!}\,t^k/2^k(k-1)!\right]. \end{aligned}$$

Now by virtue of the properties of the sequence $\{t_n\}$,

$$\sum_{k=1}^{n} k\sqrt{(2k+2)!}\,t_n^k/2^k(k-1)! = O\left[\sum_{k=1}^{n} k^3 t_n^k\right] = O[(1-t_n)^{-4}]. \tag{3.11}$$

Therefore, if $\lim h_n(1-t_n)^{-4} = 0$, we have

$$[E[\hat{G}_n(x)] - G_n(x)]^2 \sim \frac{h_n^4\, t_n^2\, \sigma^4}{36}\,[G^{(4)}(x)]^2. \tag{3.12}$$

Now consider the variance of $\hat{G}_n(x)$. We note [4] that

$$E[F_n(x)F_n(y)] = \frac{n-1}{n}\,F(x)F(y) + \frac{1}{n}\,F(\min x, y). \tag{3.13}$$

Writing out the expectation of $\hat{G}_n^2(x)$ with this substitution, we obtain

$$E[\hat{G}_n^2(x)] = \frac{n-1}{n}\,E^2[\hat{G}_n(x)] + B_n, \tag{3.14}$$

or

$$\sigma^2(\hat{G}_n(x)) = B_n - \frac{1}{n}\,E^2[\hat{G}_n(x)], \tag{3.15}$$

where

$$B_n = \frac{1}{n}\sum_{k,r=0}^{n}\left(\frac{-\sigma^2 t}{8h^2}\right)^{k+r}\frac{1}{k!\,r!}\sum_{j=0}^{2k}\sum_{s=0}^{2r}\binom{2k}{j}\binom{2r}{s}(-1)^{j+s} \cdot F[\min\,(x+(k-j)2h), (x+(r-s)2h)]. \tag{3.16}$$

For convenience in writing, let $a_n = \sigma^2 t_n/8h_n^2$. After some manipulation it can be shown that

$$B_n = \frac{1}{n}\,E[\hat{G}_n(x)] + C_n, \tag{3.17}$$

where

$$C_n = \frac{1}{n}\sum_{k,r=0}^{n}(-a_n)^{k+r}\frac{1}{k!\,r!}\sum_{s=0}^{2r}\left[\sum_{j=0}^{k-r+s-1}\binom{2k}{j}(-1)^j\right]\binom{2r}{s} \cdot(-1)^s[F(x+(r-s)2h_n) - F(x+(k-j)2h_n)]. \tag{3.18}$$

Using the fact that, for $0 \leqq m < 2k$,

$$\sum_{j=0}^{m}(-1)^j\binom{2k}{j} = (-1)^m\binom{2k-1}{m}, \tag{3.19}$$

and, from (2.9), that

$$\begin{aligned} &F(x+(r-s)2h_n) - F(x+(k-j)2h_n) \\ &\quad = 2h_n[(r-s)-(k-j)] + O[h_n^2(r-s)^2 + h_n^2(k-j)^2], \end{aligned} \tag{3.20}$$

we find that

$$C_n = \frac{2h_n}{n} \sum_{k,r=0}^{n} \frac{a_n^{k+r}}{k!\,r!} \binom{2k+2r-2}{k+r-1} + O\left[\frac{h_n^2}{n} \sum_{k,r=0}^{n} \frac{a_n^{k+r}}{k!\,r!} \binom{2k+2r-2}{k+r-1} \frac{kr}{k+r}\right]. \tag{3.21}$$

If we let $k + r = j$, (3.21) may be rewritten as

$$C_n \sim \frac{2h_n}{n} \sum_{j=1}^{2n} \frac{(2a_n)^j}{j!} \binom{2j-2}{j-1} + O\left[\frac{1}{n} \sum_{j=0}^{2n-1} \frac{(2a_n)^j}{j!} \binom{2j}{j}\right]. \tag{3.22}$$

It is easy to verify the following equalities as Taylor series:

$$\sum_{j=1}^{\infty} \frac{b^j}{j!} \binom{2j-2}{j-1} = \frac{1}{2\pi} \int_0^{\pi/2} \frac{\exp(4b\cos^2 x) - 1}{\cos^2 x}\, dx = \frac{4b}{\pi} \int_0^{\pi/2} \sin^2 x \exp(4b\cos^2 x)\, dx = bO(e^{4b}) \tag{3.23}$$

and

$$\sum_{j=0}^{\infty} \frac{b^j}{j!} \binom{2j}{j} = \frac{2}{\pi} \int_0^{\pi/2} \exp(4b\cos^2 x)\, dx = O(e^{4b}). \tag{3.24}$$

Therefore, as b increases, (3.24) becomes negligible compared with (3.23).

By the use of the integral form of the remainder,

$$\sum_{j=1}^{2n} \frac{b^j}{j!} \binom{2j-2}{j-1} = \frac{1}{2\pi} \int_0^{\pi/2} (\cos x)^{-2} \cdot \left[\frac{\exp(4b\cos^2 x)}{(2n)!} \int_{4b\cos^2 x}^{\infty} v^{2n} e^{-v}\, dv - 1\right] dx \tag{3.25}$$

and

$$\sum_{j=0}^{2n-1} \frac{b^j}{j!} \binom{2j}{j} = \frac{2}{\pi(2n-1)!} \int_0^{\pi/2} \exp(4b\cos^2 x) \left[\int_{4b\cos^2 x}^{\infty} v^{2n-1} e^{-v}\, dv\right] dx. \tag{3.26}$$

Let $b = 2a_n$. Now if $a_n/n \to 0$, it is known ([7], Chap. 7) that, for any $y \geqq 0$,

$$\lim \frac{1}{(2n)!} \int_{2a_n y}^{\infty} v^{2n} e^{-v}\, dv = 1. \tag{3.27}$$

Therefore, under this assumption, the ratios of (3.25) to (3.23), and of (3.26) to (3.24), approach unity as n increases, so that (3.26) may be neglected. As a result we have

$$C_n \sim \frac{2h_n}{n} \sum_{j=1}^{2n} \frac{(2a_n)^j}{j!} \binom{2j-2}{j-1} \tag{3.28}$$

and

$$\sigma^2(\hat{G}_n(x)) \sim \frac{E(\hat{G}_n(x))[1 - E(\hat{G}_n(x))]}{n} + \frac{2h_n}{n} \sum_{j=1}^{2n} \frac{(2a_n)^j}{j!} \binom{2j-2}{j-1}. \tag{3.29}$$

Putting together (3.29) and (3.12), and taking account of the rapidity of the convergence of $G_n(x)$ to $G(x)$, we have for the asymptotic mean square error,

$$E[(\hat{G}_n(x) - G(x))^2] \sim \frac{G(x)(1 - G(x))}{n} + \frac{h_n^4 t_n^2 \sigma^4}{36}[G^{(2)}(x)]^2 + \frac{2h_n}{n}\sum_{j=1}^{2n}\left(\frac{\sigma^2 t_n}{4h_n^2}\right)^j \frac{1}{j!}\binom{2j-2}{j-1}. \tag{3.30}$$

In order for this asymptotic expression to be valid and for the estimator to be consistent, C_n must approach zero, and h_n and t_n must obey the restrictions imposed during the derivation. These conditions on h_n and t_n are summarized here:

(A) $\lim h_n(1 - t_n)^{-4} = 0$,

and

(B) $\lim \frac{t_n}{nh_n} \exp\left[\frac{\sigma^2 t_n}{h_n^2}\right] = 0.$

Condition (B) is sufficient to ensure that C_n and a_n/n approach zero.

4. Specific sequences $\{h_n\}$ and $\{t_n\}$. The logical step, after deriving the mean square error, is to determine sequences $\{h_n\}$ and $\{t_n\}$ which minimize it. In the present case the complexity of (3.30) makes this extremely difficult. Alternatively, we may search for easily computed sequences satisfying conditions (A) and (B). Suppose we let

$$h_n = a(\ln n)^{-\alpha}, \qquad \alpha > 0 \tag{4.1}$$

and

$$t_n = 1 - (\ln n)^{-\beta}, \qquad \beta > 0. \tag{4.2}$$

Then it may be verified that if $\beta < \alpha/4$, condition (A) is satisfied. If, in addition, $\alpha = 1/2$ and $a \geqq \sigma$, then condition (B) is satisfied, and in fact C_n becomes negligible.

In order to minimize the bias, whose square is the second term of (3.30), it is reasonable to let a assume its minimum value. Finally, the convergence of G_n to G depends on the fact that $\beta > 0$. A value convenient for computation, say $\beta = 0.1$, is suggested. These "convenient" sequences are then

$$h_n = \sigma(\ln n)^{-.5} \tag{4.3}$$

and

$$t_n = 1 - (\ln n)^{-.01}. \tag{4.4}$$

5. Remarks on the bias of $\hat{G}_n(x)$. It is possible that, with smaller sample sizes, the bias introduced by the estimating procedure may be greater than the bias involved in ignoring the whole problem and simply using the original sample distribution function to estimate $G(x)$. A reasonable rule of thumb for deciding if it is worthwhile to compute $\hat{G}_n(x)$ is to do so if the maximum bias of $\hat{G}_n(x)$ does not exceed the maximum bias of $F_n(x)$ for the given sample size. To get

some idea of the order of magnitude of the sample sizes required, we will assume that $G(x)$ is a normal distribution function with variance τ^2.

The maximum bias which can result from using $F_n(x)$ is

$$\max_x | E[F_n(x)] - G(x) | = \max_{x>0} [\Phi(x/\tau) - \Phi(x/\sqrt{\sigma^2 + \tau^2})], \tag{5.1}$$

where $\Phi(x)$ is the standard normal distribution function. The maximum is attained when

$$x = \frac{\tau}{\sigma}\left[(\sigma^2 + \tau^2) \ln\left(1 + \frac{\sigma^2}{\tau^2}\right)\right]^{\frac{1}{2}}.$$

The maximum asymptotic bias of $\hat{G}_n(x)$ is

$$\max_x \frac{h_n^2 \sigma^2 t_n}{6} | G^{(4)}(x) | = \frac{h_n^2 \sigma^2 t_n}{6} \max_x | G^{(4)}(x) |, \tag{5.2}$$

where

$$G^{(4)}(x) = \frac{1}{\sqrt{2\pi}}\left(\frac{3x}{\tau^5} - \frac{x^3}{\tau^7}\right) e^{-x^2/2\tau^2}. \tag{5.3}$$

$| G^{(4)}(x) |$ attains its maximum at $x = .742\ \tau$, and

$$\max_x | G^{(4)}(x) | = \frac{1.38}{\sqrt{2\pi}\tau^4}. \tag{5.4}$$

Therefore, the maximum asymptotic bias of $\hat{G}_n(x)$ is no greater than the maximum bias of $F_n(x)$ if (using the forms (4.3) and (4.4) for h_n and t_n)

$$(\ln n)^{-1}[1 - (\ln n)^{-0.1}] \leqq 4.35\sqrt{2\pi}\left(\frac{\tau}{\sigma}\right)^4 \cdot \left[\Phi(\{(1 + \tau^2/\sigma^2) \ln (1 + \sigma^2/\tau^2)\}^{\frac{1}{2}}) - \Phi\left(\frac{\tau}{\sigma}\{\ln (1 + \sigma^2/\tau^2)\}^{\frac{1}{2}}\right)\right]. \tag{5.5}$$

Since this inequality involves the asymptotic bias of $\hat{G}_n(x)$, it is presumbly not too trustworthy for small n. Substituting $n = 30$ and solving for σ/τ, we find that (5.5) holds if $\sigma/\tau \leqq 3$. In other words, a sample of size 30 justifies computing $\hat{G}_n(x)$ if the standard deviation of the known component is no more than three times the standard deviation of the unknown component. Therefore, even though (5.5) is an asymptotic expression, it seems reasonable to say that with samples of size 30 or larger, it is worthwhile to compute $\hat{G}_n(x)$ for most situations of practical interest.

6. Dependence between Z and Y. Suppose now that the normal random variable Z, instead of having expectation zero, is dependent on Y in the sense that it has an expectation μ_y when $Y = y$. Then if $y + \mu_y$ is a continuous, strictly monotone function of y, the analyses leading to (2.1) and (3.2) are still valid, provided the derivatives and difference quotients are now evaluated at the point $x + \mu_x$. Therefore, under this kind of dependence, $\hat{G}_n(x + \mu_x)$ estimates $G(x)$.

7. Other estimators of $G(x)$. Although the present paper arose from consideration of some public health problems, the problem of instrumental error has had a long history in astronomy. In particular, several solutions to the integral equation (1.1), under varying restrictions on $G(x)$, have been given by astronomers. (See [2] and [6] for bibliographies.) With the exception of [2], however, the solutions themselves are offered as estimators, without taking into account the error involved in using an estimate of $F(x)$. Consequently, their consistency is open to question. In [2], an estimator for $G^{(1)}(x)$ when $F^{(1)}(x)$ is observed is given, and its maximum bias computed when $G^{(1)}(x)$ is a normal probability density. The approximate variance of the estimator when $F^{(1)}(x)$ is subject to error is also given, but the form of the bias shows that the estimator is not consistent.

It is of some interest to examine one of the first solutions to (1.1), given by Eddington [1]. It is

$$\sum_{k=0}^{\infty} (-1)^k F^{(2k)}(x)(\sigma^2/2)^k/k! = G(x), \tag{7.1}$$

which may be thought of as Pollard's formula (2.1) with the limit and summation operations interchanged. In practice, only the first two terms of (7.1) are used as an estimator, and difference quotients are apparently used to approximate the derivatives. However, even if we consider the whole series, and assume the derivatives known, it is clear that the convergence of the solution depends on the form of $G(x)$, so that (7.1) is not consistent for arbitrary $G(x)$.

REFERENCES

[1] A. S. Eddington, "On a formula for correcting statistics for the effects of a known probable error of observation," *Mon. Not. R. Astrom. Soc.*, Vol. 73 (1913), pp. 359–360.

[2] F. D. Kahn, "The correction of observational data for instrumental band width," *Proc. Camb. Philos. Soc.*, Vol. 51 (1955), pp. 519–525.

[3] H. Pollard, "Distribution functions containing a Gaussian factor," *Proc. Amer. Math. Soc.*, Vol. 4 (1953), pp. 578–582.

[4] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 832–837.

[5] G. Szego, *Orthogonal Polynomials*, American Mathematical Society, New York, 1939.

[6] R. J. Trumpler and H. F. Weaver, *Statistical Astronomy*, University of California Press, Berkeley, 1953.

[7] D. V. Widder, *The Laplace Transform*, Princeton University Press, Princeton, 1946.

BAYES AND MINIMAX PROCEDURES IN SAMPLING FROM FINITE AND INFINITE POPULATIONS[1]—I

By Om P. Aggarwal

Purdue University[2] and *Stanford University*

Summary. Some of the sampling methods and the methods of estimation usually employed in sample surveys are considered in terms of loss and risk functions. The loss function is taken as the sum of two components, one proportional to the square of the error of the estimate and the other proportional to the cost of obtaining the sample. Consideration is given to the problem of the allocation of the total sample size and only non-sequential estimates are discussed. As the loss function is convex and of finite expectation in each case, only non-randomized estimates are considered, since Hodges and Lehmann [5] have shown that under these conditions the class of non-randomized estimates is essentially complete. Only simple random sampling and stratified sampling methods are discussed in this part, the ratio, regression and sub-sampling methods will be discussed in subsequent parts.

1. Introduction. In the current practice of conducting sample surveys, the statisticians have adopted one of the following two procedures (see, e.g., [2], [3], [4], [7], or [8]): (i) to get an estimate of maximum precision for a given total cost of the survey, or (ii) to get an estimate of given precision for a minimum total cost of the survey. The allocation of the resources for a given survey is usually carried out, keeping in mind one or the other of the above two aims. It is possible, however, to consider jointly the losses resulting from the errors in the estimates and from the cost of sampling, and to employ such sampling and estimation procedures as will, in some sense, "minimize" the total expected loss. Accordingly we shall take as loss function the sum of two components, one proportional to the square of the error of the estimate and the other proportional to the cost of obtaining the sample. The problems generally met in sampling surveys will be formulated in terms of the decision theory using this loss function and it will be seen that their solutions are the classical results in estimation and design. This appears to be a preliminary step toward further research in this field.

2. Bayes and minimax estimates. The estimation problem with a fixed sample size has the following structure. We are given a sample space X, a space of probability distributions on X, $P_\Omega = \{p_\omega : \omega \in \Omega\}$, where Ω is an index set (generally called parameter space), and a numerical-valued function g defined on

Received April 29, 1958.

[1] Work done under the sponsorship of the Office of Naval Research while the author was at Stanford University.

[2] On leave from Purdue University during 1958–59 as Statistician with the F.A.O. of the United Nations on their Technical Assistance Mission to Chile.

Ω whose value $g(\omega)$ the statistician wishes to estimate on the basis of the outcome of an experiment, say $x \in X$. A non-randomized decision function for the statistician, usually called an estimate[3], is a numerical function δ defined on X, specifying for each x the number $a \in A$ which will be chosen to estimate $g(\omega)$ when that x is observed. The space of actions A is here the real line. The loss function L defined on $\Omega \times A$ is non-negative and is the loss incurred when $g(\omega)$ is estimated by a. The risk function R is defined by

$$R(\omega, \delta) = E_\omega L(\omega, \delta). \tag{2.1}$$

The subscript ω appended to the symbol E for expectation indicates that ω is to be regarded as fixed when expectation is taken.

If it is assumed that ω is obtained by nature as the value of a random variable having a probability distribution λ, a Bayes estimate with respect to the a priori distribution λ is defined as an estimate δ which minimizes the average risk $\int R(\omega, \delta)\, d\lambda(\omega)$. If the statistician knew λ, he would choose this estimate as his best action. But in the absence of any knowledge of λ the statistician may decide to use what is called in the terminology of two person zero-sum game a minimax strategy. A minimax estimate is defined as an estimate δ which minimizes the "maximum" risk, $\sup_{\omega \in \Omega} R(\omega, \delta)$. In the same terminology, a least favorable distribution or "maximin" strategy is defined as a distribution λ such that it maximizes the "minimum" risk $\inf_\delta \int R(\omega, \delta)\, d\lambda(\omega)$.

The following theorem [6] gives in many cases a minimax estimate as well as a least favorable distribution whenever the latter exists.

THEOREM 2.1. *If a Bayes estimate δ_λ has constant (independent of ω) risk $R(\omega, \delta_\lambda) = r$, then δ_λ is minimax and λ is a least favorable distribution.*

The following theorem [6] will give in many cases a minimax estimate where no least favorable distribution exists.

THEOREM 2.2. *If $\{\lambda_n\}$ is a sequence of a priori probability distributions, $\{r_n\}$ the sequence of associated Bayes risks, and if $r_n \to r$ as $n \to \infty$, and if there exists some estimate δ for which $R(\omega, \delta) \leqq r$ for all ω, then δ is a minimax estimate.*

We shall frequently need another theorem in the sequel. Using the term "minimax risk" for $\inf_\delta \sup_\omega R(\omega, \delta)$ we state and prove it as

THEOREM 2.3. *If δ, r are a minimax procedure and the minimax risk respectively, assuming that the observations X follow any probability distribution $\omega \in \Omega^*$, and if $\Omega \supset \Omega^*$ is a space of distributions for which the risk associated with δ does not exceed r, then δ is a minimax procedure and r the minimax risk for all distributions of X in Ω.*

PROOF. By hypothesis

$$r = \sup_{\omega \in \Omega^*} R(\omega, \delta) \leqq \sup_{\omega \in \Omega} R(\omega, \delta) \leqq r. \tag{2.2}$$

[3] It should be more properly called an estimator, to distinguish the function from a specific numerical value, but it is hoped no confusion will be caused by using the same term for both.

Hence equality holds. If d is any other procedure,

$$\sup_{\omega \in \Omega} R(\omega, \delta) = r \leqq \sup_{\omega \in \Omega^*} R(\omega, d) \leqq \sup_{\omega \in \Omega} R(\omega, d), \tag{2.3}$$

which shows that δ is a minimax procedure and r the minimax risk for all distributions in Ω.

3. Sampling from a finite population. In its simplest form, sampling from a finite population may be described in the following manner. We are given a sample space and nature (or a conscious being) performs a fixed sample size experiment and obtains a value of a random variable, which is a point $x = (x_1, x_2, \cdots, x_N)$ in the space R^N of ordered N-tuples of real numbers or vectors with real components. It should be mentioned, however, that all random variables may not be available to nature. The statistician has to select one out of a class A of possible actions in complete or partial ignorance of x and the particular probability distribution employed by nature to obtain x. He (the statistician) incurs a loss which is a bounded function of the selected action $a \varepsilon A$ and the point $x \varepsilon R^N$, but not of the underlying conceptual distribution. However, for a given a, the loss is assumed to be constant for all permutations of the coordinates of x. The statistician can obtain partial information on x by observing some fixed number of coordinates of x, say n. The problem is: If the cost of observing x_i, $i = 1, 2, \cdots, N$, is independent of i, how should the statistician select the sample and choose a?

In the case that $\Omega = R^N$ and the strategy for nature corresponding to ω is to choose $x = \omega$ with probability one, Blackwell and Girshick [1] have shown that the invariance and sufficiency principles require the sampling scheme to select each set of n distinct integers from 1 to N (without regard to order) with probability $\binom{N}{n}^{-1}$. This is the usual strategy of simple random sampling without replacement. The proof extends to the more complex situations discussed in this paper. Accordingly, we seek Bayes and minimax strategies using this sampling scheme.

4. Statement of the problem. We shall be estimating the mean of a finite population with our loss as squared error. If the variance is not restricted somehow, our risk may be arbitrarily large. Accordingly, we bound, under any strategy of nature for choosing the finite population, the expected value of the variance of the finite population. Another possibility would be to divide the loss by the expected variance. Our method actually shows that the sample mean is minimax for each expected variance.

Formally, the decision problem in which we are interested may be characterized as follows:

(a) The sample space is (X, Ω, p), where X is the N-dimensional Euclidean space R^N, Ω is the set of all distributions ω on hyperplanes in R^N of the form $x_1 + x_2 + \cdots + x_N =$ constant, say $N\mu_\omega$, and subject to the restriction that

$$E_\omega \sum_{i=1}^{N} (x_i - \mu_\omega)^2 = \int \cdots \int_{R^N} (x'x - N\mu_\omega^2)\, d\omega(x) \leqq (N-1)\sigma^2, \tag{4.1}$$

where x is the column vector with $x_1, x_2, \cdots, x_N$ as elements, σ is a given positive number, and $p_\omega = \omega$ for $\omega \in \Omega$.

(b) The action space A is the real line R^1.

(c) The loss function L, defined on $(\Omega \times A)$, is given by

$$L(\omega, a) = (a - \mu_\omega)^2. \tag{4.2}$$

(d) The space D of decision rules is the space of all ordered n-tuples of different integers from 1 through N together with all measurable mappings of n-space into A. If $\delta = (i_1, \cdots, i_n; f)$, then $\delta(x) = f(x_{i_1}, \cdots, x_{i_n})$.

The application of invariance and sufficiency principles, as in the problem stated in the last section, require the sampling scheme to select each set of ordered n-tuples of different integers from 1 through N with equal probability. Accordingly, it is sufficient to consider the following problem:

(a) The sample space (X, Ω, p), where X is the n-dimensional Euclidean space R^n, Ω is the same as before, and p_ω for $\omega \in \Omega$ is the distribution of a sample $x = (x_1, \cdots, x_n)$ obtained by simple random sampling without replacement from $x_1, x_2, \cdots, x_N$, which are distributed according to ω.

(b) The action space A is the same as before.

(c) The loss function L is the same as before.

(d) The space D of decision rules is the set of all measurable mappings of X into A.

The problem of obtaining a minimax estimate of the mean μ_ω of the finite population $(x_1, \cdots, x_N)$ is solved in the following way. Consider nature's strategy as picking μ_ω from $N(0, \theta^2)$, a normal distribution with mean zero and variance θ^2, and given μ_ω, letting ω, with probability one, be singular N-variate normal with mean μ_ω and variance $\sigma^2(N-1)/N$ for each component and covariance $-\sigma^2/N$ for each pair of components. A Bayes estimate is obtained with respect to this strategy of nature, regarded as a member of a sequence $\{\lambda_\theta\}$ of a priori distributions, and the limit, if any, of the corresponding sequence of Bayes risks $\{r_\theta\}$ as $\theta \to \infty$ is obtained, say r. Then if we can find some estimate δ for which the risk $R(\omega, \delta)$—without assuming normality of ω—does not exceed r, then by virtue of Theorems 2.2 and 2.3, δ is a minimax estimate.

The discussion so far is given in detail for the case when the sampling plan is simple random sampling without replacement. Under somewhat different situations different sampling plans will be required. We shall not discuss the derivation of the minimax strategies for the choice of sampling plans (it can be shown by an extension of Blackwell and Girshick's proof of the optimality of simple random sampling [1] that the sampling plans given are optimum under the circumstances) but will take the sampling plans as being given and confine our attention to the problem of estimating the mean of the populations by employing techniques similar to the one outlined above.

5. Bayes and minimax procedures for estimating the mean of a finite population with simple random sampling (without replacement). The average risk corresponding to an a priori distribution ξ for nature and a decision function δ

used for estimation of $g(\omega)$ by the statistician is obviously

$$R(\xi, \delta) = \underset{\Omega}{E} \underset{X}{E} [(\delta(x) - g(\omega))^2 \mid \omega], \tag{5.1}$$

where x stands for $(x_1, x_2, \cdots, x_n)$ and the symbol below E indicates the space over which the expectation is to be taken. For the sake of simplicity we shall not attempt to distinguish between a random variable and its observed value. Since the integrand is non-negative, we may change the order of taking expectations and write

$$R(\xi, \delta) = \underset{X}{E} \underset{\Omega}{E} [(\delta(x) - g(\omega))^2 \mid x]. \tag{5.2}$$

This is minimized by choosing, for each x, that number $\delta(x)$ which minimizes $\underset{\Omega}{E}[(\delta(x) - g(\omega))^2 \mid x]$ and the number $\delta(x)$ which does it is clearly

$$\delta_\xi(x) = E(g(\omega) \mid x). \tag{5.3}$$

This gives the minimum value of $\underset{\Omega}{E}[(\delta(x) - g(\omega))^2 \mid x]$ as $\sigma^2_{g(\omega)\mid x}$, the variance of the conditional distribution of $g(\omega)$ given x. Then, by (5.2), the Bayes risk r_ξ is given by

$$r_\xi = \underset{X}{E}\, \sigma^2_{g(\omega)\mid x}. \tag{5.4}$$

These results hold in general whenever the loss function for the estimation of $g(\omega)$ is of the form $L(\omega, a) = [a - g(\omega)]^2$.

In the problem under consideration, taking the a priori distribution for nature as mentioned in the last section, the distribution of the sample $(x_1, \cdots, x_n)$ given ω is n-variate normal with mean μ_ω and variance $\sigma^2(N-1)/N$ for each x_i and covariance $-\sigma^2/N$ for each pair (x_i, x_j), $i \neq j$. Since it can be shown easily that the sample mean $\bar{x}$ is a sufficient statistic for μ_ω, we see from (5.3) and (5.4) respectively that the Bayes estimate $\delta_\theta(x) = E(\mu_\omega \mid x) = E(\mu_\omega \mid \bar{x})$, and the Bayes risk $r_\theta = E\sigma^2_{\mu_\omega\mid x} = E\sigma^2_{\mu_\omega\mid \bar{x}}$. Now μ_ω is $N(0, \theta^2)$ and, given μ_ω, $\bar{x}$ is $N(\mu_\omega, v)$ where $v = (n^{-1} - N^{-1})\sigma^2$, so μ_ω and $\bar{x}$ have a bivariate normal distribution. It is then easily seen that the conditional distribution of μ_ω given $\bar{x}$ is normal with mean $\theta^2\bar{x}(\theta^2 + v)^{-1}$ and variance $\theta^2 v(\theta^2 + v)^{-1}$. Since the variance is independent of x, these are respectively Bayes estimate $\delta_\theta(x)$ and Bayes risk r_θ.

To find a minimax estimate for μ_ω, we consider if the sequence $\{r_\theta\}$ tends to a limit as $\theta \to \infty$. It is seen that it does and the limit r is given by

$$r = \lim_{\theta\to\infty} r_\theta = v = \frac{N-n}{Nn}\sigma^2. \tag{5.5}$$

By Theorem (2.2) if we can find some estimate δ for which the risk does not exceed r, then that δ is a minimax estimate. Trying $\delta(x) = \bar{x}$ $(= \lim_{\theta\to\infty} \delta_\theta(x))$,

we see that the risk corresponding to δ is given by

$$R(\omega, \delta) = E_\omega(\bar{x} - \mu_\omega)^2 = E_\omega E[(\bar{x} - \mu_\omega)^2 \mid x_1, \cdots x_N]$$

$$= \frac{N - n}{nN(N - 1)} E_\omega \sum_{i=1}^{N} (x_i - \mu_\omega)^2$$

$$\leqq \frac{N - n}{Nn} \sigma^2, \qquad \text{by (4.1)}$$

$$= r.$$

Hence $\bar{x}$ is a minimax estimate and the risk corresponding to this estimate (minimax risk) does not exceed $(N - n)\sigma^2/nN$.

6. Bayes and minimax procedures with stratified sampling plan. In survey designs stratification is a procedure whereby the entire population is divided into a number of strata and sampling is carried out independently in each stratum. Let c_i be the known cost of sampling per unit in the ith stratum, μ_i and σ_i^2 the unknown mean and the known variance of the population in the ith stratum, and let k be the number of strata. Then, for a given cost $C = \sum_{i=1}^{k} c_i n_i$, where n_i is the number of observations sampled from the ith stratum, it is well known (see e.g. [2], [3], [4], or [8]) that the procedure which estimates $\sum_{i=1}^{k} N_i \mu_i$ with minimum variance is to choose n_i proportional to $N_i\sigma_i/\sqrt{c_i}$ and to use $\sum_{i=1}^{k} N_i \bar{X}_i$ as an estimate, where N_i is the size of the ith stratum and $\bar{X}_i$ the sample mean of the n_i observations selected at random from it. The same values for n_i are obtained when for a given variance of the estimate, the object is to minimize the total cost of sampling. In this section we shall investigate for this problem some Bayes and minimax procedures, first for an infinite and then for finite populations.

A. *Infinite populations.* Suppose that the ith stratum consists of an infinite population with unknown mean μ_i and known upper bound σ_i^2 for the variance, $i = 1, 2, \cdots, k$, and that we have to estimate a linear function of the μ_i, say $U = \sum_{i=1}^{k} a_i \mu_i$ where the a_i are some given real numbers. Without loss of generality we may take $\sum a_i = 1$. For the sake of simplicity we shall assume that none of the a_i is zero. The loss function L is given by

$$L(U, \delta) = (\delta - U)^2 + \sum_{i=1}^{k} c_i n_i, \tag{6.1}$$

where $n_i(>0)$ is the size of the sample chosen from the ith stratum, c_i the sampling cost per unit in that stratum, and δ is a function of the sample $\{X_{ij}\,;\, i = 1, 2, \cdots k; j = 1, 2, \cdots, n_i\}$, where X_{ij} is the jth observation from the ith stratum. For the sake of simplicity of notation, as before, we shall not attempt to distinguish between a random variable and its observed value.

It may be noted that a slightly more realistic loss function would be

$$L(U, \delta) = \alpha(\delta - U)^2 + \sum_{i=1}^{k} c_i n_i, \tag{6.2}$$

where α is some constant depending upon the desired relative accuracy of the results and the cost of experimentation in a given situation. But it is easily seen that any procedure corresponding to this loss function can be obtained from the corresponding procedure when the loss function is (6.1) simply by substituting c_i/α for c_i. The risk associated with it will be simply α times the corresponding risk when the loss function is (6.1). For this reason also the condition $\sum a_i = 1$ in this section does not detract from the generality of the a_i.

Assume at first that, given ω, the distribution in each stratum is normal, with variance σ_i^2 in the ith stratum, $i = 1, 2, \cdots, k$. We may conjecture that for this problem there is no least favorable distribution of nature, since as U could have any real value, what we would expect it to be is a uniform distribution over the real line, but this is not a distribution.[4] We shall assume that the μ_i are normally and independently distributed each with mean zero and variance θ^2, and find Bayes solutions corresponding to the sequence $\{\lambda_\theta\}$ of a priori distributions of U resulting from the distribution of μ_i. Let δ_θ denote a corresponding Bayes estimate of U. If the Bayes risks r_θ corresponding to δ_θ tend to a limiting value r when θ tends to infinity, then any estimate which has its risk less than or equal to r will be a minimax estimate by Theorem 2.2 under the normality assumption of the observations. This assumption may then be removed easily with the help of Theorem 2.3.

We may regard the n_i as fixed for the purpose of finding the estimates. Letting δ^* be a minimax estimate for given n_i, we shall choose the n_i so as to minimize the risk,

$$R(U, \delta^*) = E(\delta^* - U)^2 + \sum_{i=1}^{k} c_i n_i, \tag{6.3}$$

as a function of the n_i.

Since we are working with fixed n_i, we shall omit the $\sum c_i n_i$ term. The loss function is now simply the square of the difference between the estimate and the quantity U being estimated and, as in the last section, Bayes estimate δ_θ and Bayes risk r_θ are given respectively by the mean and the expectation of the variance of the conditional distribution of U, given the sample x. However, since the stratum sample means are jointly sufficient for $\mu_1, \cdots, \mu_k$, we may replace the sample x in the last sentence by $\bar{X}_1, \cdots, \bar{X}_k$, where $\bar{X}_i$ is the sample mean from the ith stratum.

Now, the μ_i are independently and normally distributed with means zero and variance θ^2, and given μ_i, the $\bar{X}_i$ are independently and normally distributed with mean μ_i and variance σ_i^2/n_i, so $\mu_1, \cdots, \mu_k$ and $\bar{X}_1, \cdots, \bar{X}_k$ have a joint $2k$-variate normal distribution. It is then easily seen that the conditional distribution of μ_i, given $\bar{X}_1, \cdots, \bar{X}_k$, is normal with mean

$$y_i = \frac{n_i \theta^2 \bar{X}_i}{\sigma_i^2 + n_i \theta^2}, \tag{6.4}$$

[4] I understand from an oral communication from H. Rubin that a proof of this conjecture has been given by M. A. Girshick.

and variance

$$v_i = \frac{\theta^2\sigma_i^2}{\sigma_i^2 + n_i\theta^2}, \tag{6.5}$$

and that $\mu_1, \cdots, \mu_k$ are mutually independent given $\bar{X}_1, \cdots, \bar{X}_k$. Thus, for given $\bar{X}_1, \cdots, \bar{X}_k$, the distribution of U is normal with mean $\sum a_i y_i$ and variance $\sum a_i^2 v_i$. We thus conclude that the Bayes estimate $\delta_\theta(x) = \delta_\theta(\bar{X}_1, \cdots, \bar{X}_k) = \sum_{i=1}^k a_i y_i$, and since the variance of the conditional distribution of U is independent of $\bar{X}_1, \cdots, \bar{X}_k$, the Bayes risk $r_\theta = \sum_{i=1}^k a_i^2 v_i$.

Minimax estimate for given n_i. Letting $\theta \to \infty$, we see that $r_\theta \to r$, where $r = \sum_{i=1}^k a_i^2\sigma_i^2/n_i$. Thus, if we can find some estimate δ^* with risk $\leqq r$, then δ^* will be a minimax estimate by virtue of Theorem 2.2. Let us try the limiting Bayes estimate,

$$\lim_{\theta\to\infty} \delta_\theta(x) = \sum_{i=1}^k a_i \bar{X}_i = \delta^*(x), \quad \text{say.} \tag{6.6}$$

Since the $\bar{X}_i$ are normal and independent with means μ_i and variances σ_i^2/n_i, $\sum a_i\bar{X}_i$ is normal with mean $\sum a_i\mu_i = U$ and variance $\sum a_i^2\sigma_i^2/n_i$. Hence the risk corresponding to the estimate $\delta^*(x) = \sum a_i\bar{X}_i$ is equal to r which proves that $\sum a_i\bar{X}_i$ is a minimax estimate of U for given n_i.

It may be of interest to point out that although the Bayes estimates $\delta_\theta = \sum a_i y_i$, where y_i is given by (6.4), being unique (for given θ) are admissible, we cannot conclude from this the admissibility of δ^* because of the limiting process. The same remark applies to the other Bayes and minimax estimates obtained later, but we shall not go into the question of admissibility in this paper.

Removal of the normality assumption. Let us now do away with the assumption of normality of the distribution of X_{ij}. Suppose that whatever the joint distributions of the X_{ij}, the distribution in the ith stratum has an unknown mean μ_i, and the sample mean $\bar{X}_i$, for any sample size n_i, has a variance not exceeding a known positive number σ_i^2/n_i for $i = 1, 2, \cdots, k$, and that the $\bar{X}_i$ are uncorrelated. This is somewhat more general than the usual assumption of X_{ij} being independent with mean μ_i (unknown) and variance σ_i^2 (known) in the ith stratum, in which case the stratified sampling procedure is generally used. Let us calculate the risk R corresponding to the minimax estimate $\delta^*(x) = \sum a_i\bar{X}_i$ obtained under the assumption of the normal distribution of the observations in each stratum. It is easily verified that under these general assumptions, for given n_i,

$$\begin{aligned} R &= E\left(\sum_{i=1}^k a_i\bar{X}_i - U\right)^2 \\ &= E\left[\sum_{i=1}^k a_i(\bar{X}_i - \mu_i)\right]^2 \leqq a_i^2\sigma_i^2/n_i = r. \end{aligned} \tag{6.7}$$

Applying Theorem 2.3 now, we conclude that the minimax estimate $\sum a_i \bar{X}_i$ obtained under the assumption of normality of observations is still a minimax estimate for given n_i under the general assumptions given in the beginning of this paragraph.

Minimax strategy for choosing the n_i. Restoring the term $\sum c_i n_i$ in the risk function, one can choose "optimum" n_i if the variances of the populations in different strata are known, rather than only the upper bounds, by minimizing the risk as a function of the n_i. However, if only the upper bounds and not the actual variances are assumed to be known, the optimum choice of the n_i against the largest allowed variances σ_i^2 may not be optimum for other values of the variances, but it will still be a "minimax" choice. In other words, a minimax strategy for the statistician is to choose the n_i to be "optimum" against the maximum allowed variances σ_i^2, and then estimate U using the δ^* for these n_i. This statement follows from the following theorem.

THEOREM 6.1. *Suppose the space of strategies for the statistician is a union of spaces, say $D = \cup_c D_c$. If δ_c is minimax in D_c against Ω, the space of nature's strategies, and if $R(\omega, \delta_c)$ is constant for each c, say $R(\omega, \delta_c) = r_c$, then the δ_c minimizing r_c, if it exists, is minimax in D.*

PROOF. Let $\delta \in D$ be any strategy for the statistician. Then $\delta \in D_c$ for some c. Since δ_c is minimax in D_c,

$$\max_\omega R(\omega, \delta) \geqq \max_\omega R(\omega, \delta_c) = r_c. \tag{6.8}$$

Let δ_{c^*} be the δ_c minimizing r_c and denote the risk corresponding to δ_{c^*} by r_{c^*}. Then

$$r_c \geqq r_{c^*} = \max_\omega R(\omega, \delta_{c^*}). \tag{6.9}$$

From (6.8) and (6.9), $\max_\omega R(\omega, \delta_{c^*}) \leqq \max_\omega R(\omega, \delta)$ for all $\delta \in D$, hence δ_{c^*} is minimax in D.

We, therefore, choose optimum n_i corresponding to the variances in the different strata as the σ_i^2. For given n_i, the risk corresponding to δ^* is given by

$$R(\omega, \delta^*) = \sum_{i=1}^{k} \left[\frac{a_i^2 \sigma_i^2}{n_i} + c_i n_i \right]. \tag{6.10}$$

Now we want to choose the n_i so that this risk is minimum under the restriction that the n_i are positive integers. Since the ith term on the right hand side of (6.10) depends on n_i alone, it is sufficient to minimize $a_i^2\sigma_i^2/n_i + c_i n_i$ subject to the restriction that n_i is a positive integer. Denoting this expression by $f(n_i)$, we see that

$$f(n_i + 1) - f(n_i) = c_i - \frac{a_i^2 \sigma_i^2}{n_i(n_i + 1)}. \tag{6.11}$$

To minimize $f(n_i)$, we choose the smallest positive integral value for n_i for which the difference (6.11) is positive; in other words, the smallest positive integer n_i

for which $(n_i + 1/2)^2$ exceeds $a_i^2 \sigma_i^2/c_i + 1/4$. This gives

$$n_i = \text{integer nearest to } \sqrt{\frac{a_i^2 \sigma_i^2}{c_i} + \frac{1}{4}}. \tag{6.12}$$

When $\sqrt{a_i^2 \sigma_i^2/c_i + 1/4}$ lies exactly between two integers, say m and $m + 1$, the risk is equal and minimum for both $n_i = m$ and $n_i = m + 1$, and it is immaterial which of the two nearest integers is chosen for n_i.

B. *Finite population.* Suppose that x_{ij}, $(i = 1, 2, \cdots, k; j = 1, 2, \cdots, N_i)$ denotes some numerical characteristic of the jth unit in the ith stratum. Suppose further that the $N_i (> 1)$ are known, the means u_i of the strata are unknown, the upper bounds of the variances of the populations in the strata are known, say $\sigma_i^2 \geqq 1/(N_i - 1)E\sum_{j=1}^{N_i} (x_{ij} - u_i)^2$, and that we are required to estimate a linear function,

$$T = \sum_{i=1}^{k} a_i u_i, \tag{6.13}$$

of the population means u_i, where a_i are arbitrary known real numbers with $\sum a_i = 1$, the loss function L being given by

$$L(T, \delta) = (\delta - T)^2 + \sum_{i=1}^{k} c_i n_i, \tag{6.14}$$

where δ is an estimate for T, and c_i, n_i denote the cost of sampling per unit and the number of units sampled in the ith stratum. The sampling plan given is to decide upon k positive integers n_i, $i = 1, 2, \cdots, k$, and then choose a sample of size n_i by simple random sampling without replacement from the ith stratum, thus obtaining a sample of total size $n = \sum_{i=1}^{k} n_i$. We shall first assume that the n_i are determined somehow and obtain Bayes and minimax procedures and the corresponding risks for given n_i. Later we shall see how to choose the n_i so that the risk obtained is minimized over the choice of n_i. As before, by Theorem 6.1, this choice of the "optimum" n_i will be a minimax strategy for the statistician.

As in the case of simple random sampling discussed in Section 4, which is a special case of this problem for $k = 1$, we are considering now a decision problem in which the distribution $\omega \in \Omega$ consists of the product of k independent distributions ω_i on hyperplanes in R^{N_i} of the form $x_{i1} + \cdots + x_{iN_i} = \text{constant}$, say $N_i \mu_{\omega_i}$, and subject to the restrictions that

$$E_{\omega_i} \sum_{j=1}^{N_i} (x_{ij} - u_i)^2 \leqq (N_i - 1)\sigma_i^2, \qquad i = 1, 2, \cdots, k, \tag{6.15}$$

where the constant is denoted by $N_i \mu_{\omega_i}$ to make μ_{ω_i} the mean of $x_{i1}, x_{i2}, \cdots, x_{iN_i}$ and μ_{ω_i} itself is being written as u_i for the sake of convenience of notation, the σ_i are given positive numbers, and p_ω for $\omega \in \Omega$ is the distribution of k independent samples $x_i = (x_{i1}, \cdots, x_{in_i})$, the ith sample being obtained by simple

random sampling without replacement from $x_{i1}, \cdots, x_{iN_i}$, distributed according to ω_i.

The problem of obtaining a minimax estimate of T for given n_i is solved as before. Consider nature's strategy as picking each $u_i(=\mu_{\omega_i})$ from $N(0, \theta^2)$ and given u_i, letting the distributions ω_i, with probability one, be singular N_i-variate normal with mean u_i and variance $\sigma_i^2(N_i - 1)/N_i$ for each component and covariance $-\sigma_i^2/N_i$ for each pair of components. A Bayes estimate of U is obtained with respect to this strategy of nature, which is regarded as a member of a sequence $\{\lambda_\theta\}$ of a priori distributions, and the limit, if any, of the corresponding sequence of Bayes risks $\{r_\theta\}$ as $\theta \to \infty$ is obtained, say r. Then an estimate δ for which the risk $R(\omega, \delta)$—without assuming normality of ω—does not exceed r is, by Theorems 2.2 and 2.3, a minimax estimate for given n_i.

With nature's strategy as explained in the last paragraph, the distribution of the sample $x = \{x_{ij}\,; j = 1, \cdots, n_i, i = 1, \cdots, k\}$ given ω is the product of k distributions, the ith being n_i-variate normal with mean u_i and variance $\sigma_i^2(N_i - 1)/N_i$ for each component and covariance $-\sigma_i^2/N_i$ for each pair of components. Again, since the set $(\bar{x}_1, \cdots, \bar{x}_k)$ of the sample means from the k strata is a sufficient statistic for the set $(u_1, \cdots, u_k)$, and hence for T, we may replace the sample x in (5.3) and (5.4) by the set $(\bar{x}_1, \cdots, \bar{x}_k)$. Now, the strata are independent, i.e. the k pairs $(u_i, \bar{x}_i)$ are independent, and it follows from the calculations in Section 5 that the conditional distribution of an individual u_i given $\bar{x}_i$ is normal with mean $y_i = \theta^2\bar{x}_i(\theta^2 + v_i)^{-1}$ and variance $\theta^2 v_i(\theta^2 + v_i)^{-1}$, where $v_i = (n_i^{-1} - N_i^{-1})\sigma_i^2$. Thus, for given $\bar{x}_1, \cdots \bar{x}_k$, the conditional distribution of $T = \sum a_i u_i$ is normal with mean $\sum a_i y_i$, and variance $\sum a_i^2\theta^2 v_i(\theta^2 + v_i)^{-1}$. We thus conclude that the Bayes estimate for T is

$$\delta_\theta(x) = \sum_{i=1}^{k} a_i\, y_i = \sum_{i=1}^{k} a_i\, \theta^2\bar{x}_i(\theta^2 + v_i)^{-1}, \tag{6.16}$$

and as the variance of the conditional distribution of T is independent of x, the Bayes risk is

$$r_\theta = \sum_{i=1}^{k} a_i^2\, \theta^2 v_i(\theta^2 + v_i)^{-1} + \sum_{i=1}^{k} c_i\, n_i\,. \tag{6.17}$$

To find a minimax estimate for T now, we consider if the sequence $\{r_\theta\}$ tends to a limit as $\theta \to \infty$. It will be seen that it does, and the limit r is given by

$$\begin{aligned} r = \lim_{\theta\to\infty} r_\theta &= \sum_{i=1}^{k} a_i^2\, v_i + \sum_{i=1}^{k} c_i\, n_i \\ &= \sum_{i=1}^{k} a_i^2\, \frac{N_i - n_i}{N_i\, n_i}\, \sigma_i^2 + \sum_{i=1}^{k} c_i\, n_i\,. \end{aligned} \tag{6.18}$$

All we have to do now is to find some estimate δ^* for which the risk does not exceed r, and by Theorem 2.2, if any such δ^* exists, it will be a minimax estimate for given n_i. Trying $\delta^*(x) = \sum a_i\bar{x}_i(=\lim_{\theta\to\infty}\delta_\theta(x))$, we see that the risk cor-

responding to δ^* is given by

$$R(\omega, \delta^*) = E_\omega(\delta^* - T)^2 + \sum_{i=1}^{k} c_i n_i$$

$$= E_\omega\left[\sum_{i=1}^{k} a_i(\bar{x}_i - u_i)\right]^2 + \sum_{i=1}^{k} c_i n_i . \tag{6.19}$$

Noting the fact that the strata are independent and utilizing the result (5.6) for a single stratum, it is seen at once that (6.19) reduces to

$$R(\omega, \delta^*) \leqq \sum_{i=1}^{k} a_i^2 \frac{N_i - n_i}{N_i n_i} \sigma_i^2 + \sum_{i=1}^{k} c_i n_i = r. \tag{6.20}$$

Hence the usual estimate $\sum a_i \bar{x}_i$ is a minimax estimate for given n_i.

Minimax strategy for choosing the n_i. We now choose the n_i so that the minimax risk for given n_i and largest allowed variances in the strata is minimum under the restriction that the n_i are positive integers and $\leqq N_i$. This risk is given by

$$r = \sum_{i=1}^{k}\left[a_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)\sigma_i^2 + c_i n_i\right]. \tag{6.21}$$

This expression differs from that in (6.10) by a quantity which is independent of $n_1, n_2, \cdots, n_k$, and hence is minimized by the same n_i as before provided $n_i \leqq N_i$, and otherwise by $n_i = N_i$.

As a special case, consider the problem of the estimation of the overall mean of the finite population, $\mu = N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{N_i} x_{ij}$, where $N = \sum_{i=1}^{k} N_i$. Choosing $a_i = N_i/N$, $T = \sum a_i u_i = \mu$, and we see that a minimax procedure of estimating the mean μ is to choose the n_i as

$$n_i = \text{integer nearest to } \sqrt{\frac{N_i^2\sigma_i^2}{N^2 c_i} + \frac{1}{4}} \text{ and } \leqq N_i, \tag{6.22}$$

and then employ the usual estimate $N^{-1}\sum N_i \bar{x}_i$, for these n_i. This rule for finding the minimax n_i is more exact than the one commonly stated in the literature, namely the allocation of the total sample size in proportion to $N_i\sigma_i/\sqrt{c_i}$. This greater exactness may be quite useful in the case of high c_i's.

7. Acknowledgement. I wish to express my gratitude to the late Professor M. A. Girshick, who encouraged me to work on this problem and gave generous help and guidance. Thanks are also due to Professors David Blackwell and Herman Rubin for several helpful discussions during the course of this work. I also take this opportunity to record my appreciation of some discussions with Professor John Pratt which helped to clarify some of the concepts during the preparation of this paper.

REFERENCES

[1] David Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.

[2] W. G. Cochran, *Sampling Techniques*, John Wiley and Sons, New York, 1953.
[3] W. E. Deming, *Some Theory of Sampling*, John Wiley and Sons, New York, 1950.
[4] M. H. Hansen, W. N. Hurwitz, and W. G. Madow, *Sample Survey Methods and Theory*, Vol. II, John Wiley and Sons, New York, 1953.
[5] J. L. Hodges, Jr., and E. L. Lehmann, "Some problems in minimax point estimation," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 182–197.
[6] E. L. Lehmann, *Mimeographed notes on the Theory of Estimation*, University of California, Berkeley, Chap. IV, 1950.
[7] P. C. Mahalanobis, "On Large Scale Sample Surveys," *Phil. Trans. Roy. Soc. London, Ser. B*, Vol. 231 (1944), pp. 329–451.
[8] P. V. Sukhatme, *Sampling Theory of Surveys with Applications*, Iowa State College Press, Ames, 1954.

AN APPROXIMATION USEFUL IN UNIVARIATE STRATIFICATION

By Gunnar Ekman

Institute of Mathematical Statistics, University of Stockholm

1. Introduction. The problem of minimizing a sum $\sum_1^n P_h \sigma_h$, where P_h and σ_h^2 denote the area and conditional variance in the interval (x_{h-1}, x_h) of a density $f(x)$, arises in the theory of optimum univariate stratification (see Dalenius, [1]). In [1] Dalenius shows that the sum $\sum_1^n P_h\sigma_h$ is minimized when the conditions

$$\frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h} = \frac{\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2}{\sigma_{h+1}} \tag{1}$$

are fulfilled, $(h = 1, \cdots (n-1), x_0 = -\infty, x_n = +\infty)$, where μ_h denotes the conditional mean in the interval (x_{h-1}, x_h).

In order to avoid the computational difficulties presented by determining $\{x_h\}$ such that the conditions (1) are satisfied, various approximations to (1) have been proposed. A brief summary of these results is given in [2]. In this article a new approximation will be derived and numerical examples will be given.

We shall show, that under certain conditions[1] and for a density over a finite range, points $\{x_h\}$ satisfying the equalities

$$(x_h - x_{h-1})P_h = C_n, \qquad h = 1, 2, \cdots, n. \tag{1}^{(a)}$$

where C_n is a constant dependent on n, approximately satisfy the minimal conditions (1). For a density over an infinite range the above is obviously not applicable, in which case, however, a certain modification can be made, substituting the conditions (12), (13) below for $(1)^{(a)}$. The basic result will be derived under the assumption of a large n, i.e. correspondingly small intervals $(x_h - x_{h-1})$; asymptotically as n approaches infinity the conditions (1) and $(1)^{(a)}$ will be proven equivalent. In practice, of course, n is often rather small, hardly greater than 4 or 5, which certainly does not fulfill this requirement of a large n. It is still possible that even for $n = 2$, 3, etc., the points satisfying $(1)^{(a)}$ provide a good approximation to (1) in the sense of a $\sum_1^n P_h\sigma_h$ near the minimal value. In order to ascertain whether this may be the case or not, actual numerical computation has been carried out for three densities and for $n = 2$, 3, 4, and 5. In the table under "Numerical Examples" the points $\{x_h\}$ obtained by applying $(1)^{(a)}$ or the substitute conditions (12), (13), may be

Received February 28, 1958; revised July 3, 1958.

[1] These conditions, which concern primarily the regularity of the density function $f(x)$, are imposed in order to facilitate the mathematical derivation of the final result, and should be of no interest or concern to the reader desirous only of applying this result in practice.

compared with the points satisfying the actual minimal conditions (1); below is given a comparison of the respective $[\sum P_k \sigma_k]^2$:

n	$f(x) = 2(1-x)$		$f(x) = e^{-x}$		$f(x) = xe^{-x}$	
	$(1)^{(a)}$	(1)	$(1)^{(a)}$	(1)	$(1)^{(a)}$	(1)
2	0.01517	0.01505	0.2856	0.2855	0.6420	0.6370
3	0.00693	0.00688	0.1333	0.1332	0.3090	0.3075
4	0.00395	0.00393	0.0769	0.0768	0.1811	0.1804
5	0.00255	0.00254	0.0500	0.0500	0.1189	0.1185

It therefore seems not unlikely that the conditions $(1)^{(a)}$ are suitable substitutes for the exact conditions (1); the differences in the above table are seen to be comparatively small and decrease both absolutely and percentagewise as n increases. It is nevertheless appropriate to mention a general type of density for which $(1)^{(a)}$ may be expected to give rather poor results, namely such $f(x)$ with a long (finite) "tail" (for example a χ^2-distribution with a large number of d.f.) or with $f(x) = 0$ in an end point. In this case it is advisable to use (12), (13) unless these are difficult analytical expressions; in the case of an infinite "tail" this is imperative.

The determination of points $\{x_k\}$ satisfying $(1)^{(a)}$ or (12), (13) is by no means entirely free of computational difficulties, although these points are of course considerably easier to find than those satisfying (1). Trial and error or various iterative methods may suffice in some cases, but the need for a more systematic approach arises even for comparatively convenient expressions for $f(x)$. It is apparent that with a knowledge of the constant C_n the determination of the points becomes trivial, as they may then be found simply by "fitting" with the aid of a table over the distribution function. An iterative method of finding C_n is described in the paragraphs "First approximation to the constant C_n" and "Adaption to Numerical Computation". The mathematical arguments leading to $(1)^{(a)}$ following in the next paragraph are rather simple and straightforward, involving only the use of Taylor series and elementary algebraic manipulation. An outline of a method of finding the points satisfying the exact minimal conditions (1) concludes the paper.

2. Approximation over a finite range. Consider a density $f(x)$ over a finite range (x_0, x_n). We assume that the derivatives $f'(x)$ and $f''(x)$ exist and are continuous over the whole range. We introduce a function $H(x)$ by defining

$$H'''(x) = f(x),$$

$$H''(x) = \int_{-\infty}^{x} H'''(t)\, dt,$$

$$H'(x) = \int_{-\infty}^{x} H''(t)\, dt,$$

$$H(x) = \int_{-\infty}^{x} H'(t)\, dt.$$

$H'(x) \cdots H^{(5)}(x)$ exist therefore and are continuous. One finds by partial integration:

$$\int_a^b x^2 f(x)\,dx = 2[H(b) - H(a)] - 2[bH'(b) - aH'(a)] + [b^2H''(b) - a^2H''(a)],$$

$$\int_a^b xf(x)\,dx = -[H'(b) - H'(a)] + [bH''(b) - aH''(a)],$$

$$\int_a^b f(x)\,dx = [H''(b) - H''(a)].$$

By substituting the expressions in the right hand members of these equalities in

$$P_h\,\mu_h = \int_{x_{h-1}}^{x_h} xf(x)\,dx,$$

$$P_h[\sigma_h^2 + \mu_h^2] = \int_{x_{h-1}}^{x_h} x^2 f(x)\,dx,$$

where P_h, μ_h, and σ_h are the area, conditional mean and conditional variance of $f(x)$ in the interval (x_{h-1}, x_h), we obtain by some calculations:[2]

$$P_h[\sigma_h^2 + (x_h - \mu_h)^2] = 2\left\{H(x_h) - \left[H(x_{h-1}) + (x_h - x_{h-1})H'(x_{h-1}) + \frac{(x_h - x_{h-1})^2}{2!}H''(x_{h-1})\right]\right\},$$

$$P_h[x_h - \mu_h] = \{H'(x_h) - [H'(x_{h-1}) + (x_h - x_{h-1})H''(x_{h-1})]\},$$

$$P_h = [H''(x_h) - H''(x_{h-1})].$$

We note that the right hand members consist of $H(x_h)$, $H'(x_h)$ and $H''(x_h)$ minus the first terms in their respective Taylor expansions about the point $x = x_{h-1}$. Continuing these expansions, the following results are obtained:

$$P_h[\sigma_h^2 + (x_h - \mu_h)^2] = 2\left[\frac{(x_h - x_{h-1})^3}{3!}H'''(x_{h-1}) + \frac{(x_h - x_{h-1})^4}{4!}H^{(4)}(x_{h-1}) + \frac{(x_h - x_{h-1})^5}{5!}H^{(5)}(\xi_1)\right], \tag{2}$$

$$P_h[x_h - \mu_h] = \left[\frac{(x_h - x_{h-1})^2}{2!}H'''(x_{h-1}) + \frac{(x_h - x_{h-1})^3}{3!}H^{(4)}(x_{h-1}) + \frac{(x_h - x_{h-1})^4}{4!}H^{(5)}(\xi_2)\right], \tag{3}$$

$$P_h = \left[\frac{(x_h - x_{h-1})}{1!}H'''(x_{h-1}) + \frac{(x_h - x_{h-1})^2}{2!}H^{(4)}(x_{h-1}) + \frac{(x_h - x_{h-1})^3}{3!}H^{(5)}(\xi_3)\right], \tag{4}$$

[2] These identities may also be obtained by partially integrating $\int_{x_{h-1}}^{x_h} (x_h - x)^k f(x)\,dx$, $k = 0, 1, 2$.

where ξ_i are points in the interval (x_{h-1}, x_h). By multiplying (2) by (4) and subtracting the square of (3), the following identity is obtained:

$$(5)\quad [P_h\,\sigma_h]^2 = \frac{(x_h - x_{h-1})^4}{12}\{H'''(x_{h-1})[H'''(x_{h-1}) + (x_h - x_{h-1})H^{(4)}(x_{h-1})] + R_1\}.$$

Squaring both members of (2) we obtain:

$$(6)\quad P_h^2[\sigma_h^2 + (x_h - \mu_h)^2]^2 = \frac{(x_h - x_{h-1})^6}{9}\left\{H'''(x_{h-1})[H'''(x_{h-1}) + \frac{H^{(4)}(x_{h-1})}{2!}(x_h - x_{h-1})] + R_2\right\}.$$

R_1 and R_2 are terms of the second order or higher in $(x_h - x_{h-1})$. For large n, that is for small intervals $(x_h - x_{h-1})$, these terms may be neglected, introducing thereby only a slight error. Substituting $f(x)$ and $f'(x)$ for $H'''(x)$ and $H^{(4)}(x)$, and dividing (6) by (5), the following approximative identity is obtained:

$$(7)\quad \left[\frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h}\right]^2 \sim \frac{4(x_h - x_{h-1})}{3} \cdot \frac{f(x_{h-1})(x_h - x_{h-1}) + \frac{f'(x_{h-1})}{2!}(x_h - x_{h-1})^2}{f(x_{h-1}) + f'(x_{h-1})(x_h - x_{h-1})}.$$

In the numerator of the second factor of the right hand member of (7) we have the first two terms of the Taylor expansion of P_h about the point $x = x_{h-1}$, whereas the denominator is the partial derivative of the numerator with respect to x_h, i.e. the first two terms of the expansion of $f(x_h)$ in the same point. Approximating once again by neglecting terms of greater order in the expansions of P_h and $f(x_h)$, we obtain finally

$$(8)\quad \left[\frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h}\right]^2 \sim \frac{4(x_h - x_{h-1})P_h}{3f(x_h)}.$$

Proceeding in entirely the same fashion the analogous expression

$$(9)\quad \left[\frac{\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2}{\sigma_{h+1}}\right]^2 \sim \frac{4(x_{h+1} - x_h)P_{h+1}}{3f(x_h)}$$

may be derived.

Applying these results to the identity (1), one obtains

$$(10)\quad (x_h - x_{h-1})P_h \sim (x_{h+1} - x_h)P_{h+1}.$$

Applying (10) finally with $h = 1, \cdots, (n-1)$, we find an approximation to (1), namely

$$(11)\quad (x_h - x_{h-1})P_h = C_n,$$

where C_n is a constant.

We have tacitly assumed $f(x) \neq 0$; the case $f(x) = 0$ will usually present itself only for $x = x_0$ or $x = x_n$, in which case (12), (13) below may be used instead of (11). When $f(x) \neq 0$, and assuming without loss of generality a range of unit length, we see from (7) that the first neglected term in the numerator is $O(x_h - x_{h-1})^4$, which implies the same degree of approximation in (8), (9). Accordingly the square roots of the members of (8), (9) differ by terms of the third order. This in turn implies both that the partial derivatives of $\sum P_h\sigma_h$ (being in fact $(f(x_h)/2)\{[\sigma_h^2 + (x_h - \mu_h)^2]/\sigma_h - [\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2]/\sigma_{h+1}\}$) are $O(x_h - x_{h-1})^3 + O(x_{h+1} - x_h)^3$ in the points satisfying (11), and that the approximate $\{x_h\}$ must be adjusted by $\sum_1^n O(x_h - x_{h-1})^3 = O(x_j - x_{j-1})^2 = K_h$ in order to satisfy (1), since the members of (1) are of the first order. We deduce from these results and from the expansion to $\sum (a_hK_h^2 + b_hK_hK_{h+1})$ of $\sum P_h\sigma_h$ about the points satisfying (11) that the approximate and true minimal values of the sum differ by a sum of n terms each $= O(x_j - x_{j-1})^4$, that is by $O(x_j - x_{j-1})^3$. The conditions (11) thus generate points $\{x_h\}$ differing from the true minimal points by terms of greater order than the interval lengths and result in an approximate minimal $\sum P_h\sigma_h$ differing from the true value by a term of greater order than the sum itself, i.e., $O(x_j - x_{j-1})$; these differences should furthermore decrease monotonically as n increases, that is as $(x_j - x_{j-1})$ decreases. These conclusions are immediately extended to comprise even the case $f(x) = 0$ in the end points, and are therefore valid for any reasonable density over a finite range. By a truncation argument we may immediately ascertain the asymptotic equivalence of the sums, when (12), (13) below are used to approximate (1), even over an infinite range, whereas in this case the two sets of points do not necessarily converge, as may be seen by taking $f(x) = e^{-x}$.

3. Approximation over an infinite range. If $x_0 = -\infty$, $x_n = +\infty$, the approximation (11) can still be applied for $h = 2, \cdots, (n-1)$. The identities (8) and (9) with $h = 1$ and $(n-1)$ respectively suggest putting

$$(12) \qquad (x_h - x_{h-1})P_h = C_n, \qquad h = 2, \cdots, (n-1),$$

$$(13) \qquad \frac{3f(x_1)}{4}\left[\frac{\sigma_1^2 + (x_1 - \mu_1)^2}{\sigma_1}\right]^2 = \frac{3f(x_{n-1})}{4}\left[\frac{\sigma_n^2 + (x_{n-1} - \mu_n)^2}{\sigma_n}\right]^2 = C_n,$$

whereby an approximation over the infinite range is obtained. The functions in the left hand members of (13) depend only on the variables x_1 and x_{n-1} respectively and are often convenient analytical expressions, e.g. for $f(x) = e^{-x}$,

$$\frac{3f(x_{n-1})}{4}\left[\frac{\sigma_n^2 + (x_{n-1} - \mu_n)^2}{\sigma_n}\right]^2 = 3e^{-x_{n-1}}.$$

The result (11) may be compared to some analogous results discussed in [2]. In [5] it is proven that $P_h\sigma_h = C_n$ gives an approximate solution to (1). Now under the assumption of small intervals $(x_h - x_{h-1})$, $f(x)$ may be approximated by a constant $f(\xi)$ in this interval $(x_{h-1} \leqq \xi \leqq x_h)$, in which case

$$\sigma_h = (x_h - x_{h-1})/\sqrt{12}.$$

Therefore this result [5] and (11) are substantially equivalent, each one follows from the other. (11) might seem advantageous from a computational point of view. From a theoretical point of view we have at the same time derived (8) and (9), which permits (12) and (13) to be used in some cases mentioned in the introduction, where both (11) and the result in [5] give comparatively poor results.

In [4] it is shown that $[x_k - x_{k-1}] = C_n$ also gives $\{x_n\}$ approximating (1). This result follows from (7) above, if there in the series expansions we neglect even the second terms, obtaining then

$$\frac{\sigma_k^2 + (x_k - \mu_k)^2}{\sigma_k} \sim \frac{2}{\sqrt{3}}(x_k - x_{k-1}),$$

and the analogous result for the right hand member of (1). It is therefore reasonable to assume that Aoyama's result gives a poorer approximation (at least this is so asymptotically) than (11) or the result in [5], although computationally, of course, $(x_k - x_{k-1}) = C_n$ is better than both of these; it has nevertheless the disadvantage of being restricted to a finite range.

The approximation $P_k\mu_k = C_n$ has been proposed. We see from (11) that asymptotically this would imply $(x_k - x_{k-1})/\mu_k =$ a constant, so that the interval lengths would increase with k, whereas the P_k necessarily decrease. This so-called principle of equipartition might therefore be of use when dealing with decreasing $f(x)$, e.g. of exponential type.

4. First approximation of the constant C_n. With a knowledge of C_n the set $\{x_k\}$ satisfying (11) or (12), (13) may easily be found, which set then approximately satisfies (1). The necessity of at least an approximation to C_n arises.

We shall derive an approximation of C_n under the assumption that C_{n-1} has already been obtained; this will be done by the following heuristic argument. Suppose that for a density with a finite range $(x_n - x_0)$ a set of points $x_1, \cdots, x_{n-1}$ has been found such that the relations (11) are satisfied (there is always a unique such solution, as can be seen immediately). The left side of the identity

$$\sum_{k=1}^{n}\left(\frac{x_k - x_{k-1}}{x_n - x_0}\right)P_k = \frac{nC_n}{(x_n - x_0)} \tag{14}$$

may be considered as a weighted mean either of the $(x_k - x_{k-1})/(x_n - x_0)$ weighted by the P_k or vice versa. We noticed above that Aoyama's approximation $(x_k - x_{k-1}) = C =$ a constant is asymptotically correct when all terms but the first are neglected in (5) and (6), so that for large n

$$[(x_k - x_{k-1})/(x_n - x_0)] \sim 1/n,$$

and (14) becomes

$$\frac{1}{n}\sum_{k=1}^{n}P_k = \frac{1}{n} \sim \frac{nC_n}{(x_n - x_0)},$$

that is:

$$C_n \sim \frac{(x_n - x_0)}{n^2}. \tag{15}$$

As a first approximation C'_n to C_n the expression

$$C'_n = \frac{(n-r)^2 C_{n-r}}{n^2} \tag{16}$$

may therefore be used.

Of course other first approximations C'_n might be conceived of. There is for example some logic in adjusting C'_n as given by (16) by the factor C_{n-r}/C'_{n-r}, thereby obtaining

$$C'_n = \frac{(n-r)^2 C_{n-r}^2}{n^2 C'_{n-r}}. \tag{17}$$

5. Adaption to numerical computation. The above results (11) and (12), (13), together with (16), (17) suggest a reasonably simple method of finding a set of points $\{x_h\}$ which approximatively satisfy the identities (1), that is, which approximately minimize the sum $\sum_1^n P_h\sigma_h$ for a given density. The method to be described is applicable to the case of a finite range; in the general case the identities (13) are used instead of $(x_1 - x_0)P_1 = (x_n - x_{n-1})P_n = C_n$.

Assuming that the set $\{x_h\}$ satisfying (11) has been found for some $n = n - r$, (e.g. for $n = 2$, x_1 may be found by trial and error), a first approximation to C_n is obtained by (16) or (17). A set $\{x'_h\}$ may be found such that (in general) all $(x'_h - x'_{h-1})P'_h$ but one are equal to C'_n. Let us assume that $(x'_j - x'_{j-1})P'_j \neq C'_n$ (e.g. $j = n$). Then a second approximation C''_n to C_n may be obtained by putting

$$C''_n = \frac{(n-1)C'_n + (x'_j - x'_{j-1})P'_j}{n}. \tag{18}$$

Proceeding in the same manner a set $\{x''_h\}$ such that all $(x''_h - x''_{h-1})P''_h$ but one are equal to C''_n may be found, and a new approximation C'''_n to C_n is obtained by an analogous formula to (18), etc. The $C_n^{(i)}$ thus obtained converge to C_n, and the sets $\{x_h^{(i)}\}$ correspondingly to the set $\{x_h\}$ satisfying (11).

6. Numerical examples. The method described above has been applied to three densities, whereof two are over a semi-finite range. For $n = 2$ the point x_1 has been found by trial and error. As first approximations C'_n, (16) with $r = 1$, has been used in all cases but one ($f(x) = xe^{-x}$, $n = 4$), where (17) with $r = 1$ was used. The points $\{x_h\}$ have been found to three decimals; as a rule a comparable degree of accuracy will not be necessary, at least for the first approximation.

The results are summarized in table I. The results under Min. are the points satisfying the exact minimal conditions (1), with reservation for the last decimal, which may differ by a unit from the true value. The exact minimal variances are given in the last column.

We note that both $f(x) = 2(1 - x)$ and $f(x) = xe^{-x}$ are densities of the type mentioned in the introduction as being not quite suitable for application of equations (1)[a]; here better approximations can be obtained by using (13) for $h = n$ in the first, $h = 1$ (and of course $h = n$) in the second case.

7. Note on the method used in obtaining the exact minimal points. The method described above gives in many cases a fairly good approximate solution

TABLE I

Number of intervals	Approximation	x_1	x_2	x_3	x_4	x_5	C	$(\Sigma P_{\nu n h})^2$	
2	1	0.382	1.000					0.01517	
	min.	0.354	1.000				$C_2 = 0.2361$	0.01505	
3	1	0.244	0.528	1.000			$C_3' = 0.1049$	0.00692	
	2	0.242	0.531	1.000			$C_3'' = 0.1030$	0.00693	
	min.	0.230	0.503	1.000			$C_3''' = 0.1029$	0.00688	$f(x)$
4	1	0.178	0.379	0.613	1.000		$C_4' = 0.0579$	0.00395	$= 2(1$
	2	0.177	0.376	0.615	1.000		$C_4'' = 0.0573$	0.00395	$- x)$
	min.	0.171	0.361	0.588	1.000		$C_4''' = 0.0573$	0.00393	
5	1	0.141	0.294	0.466	0.668	1.000	$C_5' = 0.0367$	0.00255	
	2	0.140	0.292	0.463	0.669	1.000	$C_5'' = 0.0364$	0.00255	
	min.	0.136	0.283	0.448	0.644	1.000	$C_5''' = 0.0365$	0.00254	
2	1	1.233	∞					0.2856	
	min.	1.262	∞				$C_2 = 0.8737$	0.2855	
3	1	0.742	2.045	∞			$C_3' = 0.3883$	0.1334	
	2	0.766	1.991	∞			$C_3'' = 0.4095$	0.1333	
	3	0.763	1.997	∞			$C_3''' = 0.4071$	0.1333	
	min.	0.764	2.026	∞			$C_3'''' = 0.4071$	0.1332	
4	1	0.545	1.295	2.572	∞		$C_4' = 0.2292$	0.0769	$f(x) =$
	2	0.553	1.317	2.547	∞		$C_4'' = 0.2349$	0.0769	e^{-x}
	min.	0.551	1.315	2.577	∞		$C_4''' = 0.2345$	0.0768	
5	1	0.430	0.977	1.730	2.995	∞	$C_5' = 0.1501$	0.0500	
	2	0.433	0.985	1.748	2.982	∞	$C_5'' = 0.1522$	0.0500	
	min.	0.431	0.982	1.746	3.008	∞	$C_5''' = 0.1521$	0.0500	
2	1	2.125	∞					0.6420	
	min.	2.291	∞				$C_2 = 1.3319$	0.6370	
3	1	1.423	2.976	∞			$C_3' = 0.5920$	0.3125	
	2	1.467	3.109	∞			$C_3'' = 0.6328$	0.3092	
	3	1.472	3.123	∞			$C_3''' = 0.6368$	0.3090	
	min.	1.571	3.252	∞			$C_3'''' = 0.6371$	0.3075	
4	1	1.175	2.304	3.949	∞		$C_4' = 0.3859$	0.1809	$f(x) =$
	2	1.157	2.259	3.835	∞		$C_4'' = 0.3720$	0.1811	xe^{-x}
	min.	1.234	2.324	3.915	∞		$C_4''' = 0.3696$	0.1804	
5	1	0.955	1.785	2.792	4.277	∞	$C_5' = 0.2365$	0.1191	
	2	0.960	1.796	2.814	4.327	∞	$C_5'' = 0.2396$	0.1189	
	3	0.961	1.798	2.818	4.336	∞	$C_5''' = 0.2401$	0.1189	
	min.	1.032	1.859	2.876	4.425	∞	$C_5'''' = 0.2401$	0.1185	

to the problem of minimizing $\sum P_h\sigma_h$, from the point of view of generating points $\{x_h\}$ fairly close to the true minimal values. When this is the case, the "usual" iterative procedure of finding successively better approximations converging to the true values may be employed, which method will be briefly reviewed for completeness.

To this end denote by A_h the left hand member of (1), B_{h+1} the right hand member, that is:

$$A_h = \frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h},$$

$$B_{h+1} = \frac{\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2}{\sigma_{h+1}}.$$

The minimal conditions (1) then take on the form

$$A_h - B_{h+1} = 0, \qquad h = 1, \cdots, (n-1).$$

The conditions $(1)^{(a)}$ or (12), (13) have resulted in approximative values of these expressions, which may be denoted by $A_h^{(0)}$, $B_{h+1}^{(0)}$.

Taylor expansions of A_h and B_{h+1} about $x_{h-1}^{(0)}$, $x_h^{(0)}$, and $x_h^{(0)}$, $x_{h+1}^{(0)}$, respectively, have then the form

$$A_h\,(x_{h-1}^{(0)} + K_{h-1}, x_h^{(0)} + K_h) = A_h^{(0)} + \left(\frac{\partial A_h}{\partial x_{h-1}}\right)_0 K_{h-1} + \left(\frac{\partial A_h}{\partial x_h}\right)_0 K_h,$$

$$B_{h+1}(x_h^{(0)} + K_h, x_{h+1}^{(0)} + K_{h+1}) = B_{h+1}^{(0)} + \left(\frac{\partial B_{h+1}}{\partial x_h}\right)_0 K_h + \left(\frac{\partial B_{h+1}}{\partial x_{h+1}}\right)_0 K_{h+1},$$

if terms of second order or higher are neglected, and where the subscript 0 denotes the value of the partial derivatives in $x_h^{(0)}$, etc. We should now attempt to find $\{K_h\}$ so that these expressions equalize, that is, we should solve the system of linear equations:

(a)
$$\left(\frac{\partial A_h}{\partial x_{h-1}}\right)_0 K_{h-1} + \left[\left(\frac{\partial A_h}{\partial x_h}\right)_0 - \left(\frac{\partial B_{h+1}}{\partial x_h}\right)_0\right] K_h - \left(\frac{\partial B_{h+1}}{\partial x_{h+1}}\right)_0 K_{h+1}$$
$$= B_{h+1}^{(0)} - A_h^{(0)}, \qquad h = 1, \cdots, (n-1).$$

whereafter K_h is added to $x_h^{(0)}$ to obtain the new approximative x_h. The matrix of this equation will be >0 if the first approximation is good, since this matrix is definite positive in the exact minimal points. The system may be solved in a relatively simple manner by first finding for example K_1 by Cramer's formula, and by then successively determining K_2, K_3, $\cdots$ from the equations for $h = 1, 2, \cdots$. It may be noted that the matrix is a so called continuant matrix, and has a simple recursive form discussed in the literature. Equations (a) are of course used iteratively, (if necessary), once the first set of $\{K_h\}$ has been found, etc.

It remains to find the expressions for the partial derivatives appearing in the equations (a). We first multiply both numerator and denominator of A_k by P_k. From the expressions for $P_k[\sigma_k^2 + (x_k - \mu_k)^2]$ and $P_k[x_k - \mu_k]$ in the function H and its derivatives, we find immediately

$$\frac{\partial}{\partial x_k}[P_k[\sigma_k^2 + (x_k - \mu_k)^2]] = 2P_k[x_k - \mu_k],$$

and by simple derivation

$$\frac{\partial}{\partial x_{k-1}}[P_k[\sigma_k^2 + (x_k - \mu_k)^2]] = -(x_k - x_{k-1})^2 f(x_{k-1}).$$

Keeping in mind that the original minimal conditions were derived from

$$\frac{\partial[P_k \sigma_k]}{\partial x_{k-1}} = -\frac{f(x_{k-1})}{2} B_k,$$

$$\frac{\partial[P_k \sigma_k]}{\partial x_k} = \frac{f(x_k)}{2} A_k,$$

we then find by simple computation (the procedure is analogous for B_{k+1}):

$$\text{(b)}\quad \begin{aligned} \frac{\partial A_k}{\partial x_{k-1}} &= \frac{f(x_{k-1})}{P_k \sigma_k}\left[\frac{A_k B_k}{2} - (x_k - x_{k-1})^2\right], \\ \frac{\partial A_k}{\partial x_k} &= \frac{1}{P_k \sigma_k}\left[2P_k(x_k - \mu_k) - \frac{f(x_k)}{2} A_k^2\right], \\ \frac{\partial B_{k+1}}{\partial x_k} &= \frac{1}{P_{k+1} \sigma_{k+1}}\left[2 P_{k+1}(x_k - \mu_{k+1}) + \frac{f(x_k)}{2} B_{k+1}^2\right], \\ \frac{\partial B_{k+1}}{\partial x_{k+1}} &= \frac{f(x_{k+1})}{P_{k+1} \sigma_{k+1}}\left[(x_{k+1} - x_k)^2 - \frac{A_{k+1} B_{k+1}}{2}\right] = -\frac{f(x_{k+1})}{f(x_k)} \cdot \frac{\partial A_{k+1}}{\partial x_k}. \end{aligned}$$

These expressions involve only the functions A_k, etc. themselves, and μ_k, σ_k, $f(x_k)$, etc., which have already been calculated to obtain A_k, etc., and are therefore not difficult to compute.

Using equations (a) on the approximate values found by $(1)^{(a)}$ for the three densities above, the exact values (with reservation for the third decimal in a few cases) were found with only one application of (a). This would seem to imply, that no greater error is introduced by neglecting second order terms and higher in the expansions of A_k, B_{k+1}, and that application of the above method on the points satisfying conditions $(1)^{(a)}$ very often provides a reasonably simple and systematic method of finding points satisfying the minimal conditions (1).

Acknowledgements. I stand in appreciation to Docent Tore Dalenius of the Institute of Statistics, University of Stockholm for stating the problem and allowing me to examine his article [2] before it appeared in print.

I am also indebted to the referee and to Mr. J. M. Hammersley for helpful suggestions in the revision of the original paper.

REFERENCES

[1] T. DALENIUS, "The problem of optimum stratification," *Skand. Aktuarietids.*, (1950), pp. 203–213.

[2] T. DALENIUS AND J. L. HODGES, JR. "Minimum variance stratification," *J. Amer. Stat. Assn.*, (1958), (submitted in December 1957).

[3] T. DALENIUS AND M. GURNEY, "The problem of optimum stratification II," *Skand. Aktuarietids.*, (1951), pp. 133–148.

[4] H. AOYAMA, "A study of the stratified random sampling," *Ann. Inst. Stat. Math.*, (1954), pp. 1–36.

[5] T. DALENIUS AND J. L. HODGES, JR. "The choice of stratification points," (1958), (Soon to appear in *Skand. Aktuarietids.*).

TRUNCATION AND TESTS OF HYPOTHESES[1]

By Om P. Aggarwal[2,3] and Irwin Guttman[4]

Purdue University and Princeton University

1. Summary. This paper examines the loss of power when using tests based on the assumption that the variable being sampled has a "complete" normal distribution when in fact the distribution is a "truncated" one. The cases considered here are for small sample sizes and "symmetric" truncation, while the hypothesis considered is the one-sided testing for the mean of a normal distribution. Some tables are computed and it appears that an appreciable loss occurs only in the size of the test. The loss in power is found to decrease very rapidly with the distance of the alternative value of the mean from the one tested and also with the distance of the truncation from the mean.

2. Introduction. In sampling from a normal distribution the assumption that the random variable X is defined over $(-\infty, \infty)$ is an unrealistic one, and "a sample of n from a normal distribution" is in reality a sample of n from a "truncated" normal distribution. This problem has been dealt with from various points of view in several recent papers (see references). However, one aspect that seems to have been neglected is that of the tests of hypotheses. We shall attempt to examine the results of applying some usual tests of hypotheses to the case when the available sample is known to have come from a truncated population.

We call a normal distribution 'symmetrically truncated' at the 'terminus point' a if its density is given by

$$f(x) = \frac{c}{\sqrt{2\pi}\sigma} \exp\left[-\tfrac{1}{2}(x-\mu)^2/\sigma^2\right], \quad \text{for } |x-\mu| < a\sigma, \tag{2.1}$$

$$= 0 \qquad \text{otherwise,}$$

where c is given by

$$\frac{1}{c} = \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} e^{-t^2/2}\, dt. \tag{2.2}$$

We shall confine our attention to the problems of symmetric truncation only, with a and σ^2 known.

Received January 14, 1958; revised June 30, 1958.

[1] Research commenced while both authors were at the University of Alberta, under grant from the University of Alberta General Research Fund.

[2] Research continued at the Summer Research Institute of the Canadian Mathematical Congress at Kingston.

[3] At present on leave from Purdue University and with the F.A.O. of the United Nations in Chile.

[4] Research continued on support by the Office of Naval Research under Contract Nonr-1858 (05).

3. Distribution of sample means. Suppose a sample $X_1, \cdots, X_n$ of size n is available from a distribution of the form (2.1). The sampling distribution of $\bar{X} = 1/n\sum_{i=1}^{n} X_i$ for arbitrary n is very complicated and no general formula giving the distribution of $\bar{X}$ explicitly is available. However, by using convolutions of distributions, it is quite easy to derive the distribution of $\bar{X}$ for small values of n. The results for $n = 1, 2, 3$ and 4 are given below where without loss of generality $\mu = 0$, $\sigma = 1$. The density function of $\bar{X}$ is denoted by $f_n(x)$.

Case $n = 1$. From (2.1) the density is given by

$$f_1(x) = \begin{cases} \dfrac{c}{\sqrt{2\pi}} \exp(-x^2/2), & \text{for } |x| < a, \\ 0 \quad \text{otherwise,} \end{cases} \tag{3.1}$$

where c is given by (2.2).

Case $n = 2$. Using convolution on (3.1) we obtain

$$f_2(x) = \begin{cases} \dfrac{\sqrt{2}c^2}{\pi} e^{-x^2} \displaystyle\int_0^{\sqrt{2}(a-|x|)} e^{-t^2/2}\, dt, & \text{for } |x| < a, \\ 0 \quad \text{otherwise.} \end{cases} \tag{3.2}$$

Case $n = 3$. Convoluting (3.1) and (3.2) it can be verified that

$$f_3(x) = \begin{cases} \dfrac{\sqrt{6}c^3}{2\pi^{3/2}} e^{-(3/2)x^2} \displaystyle\int_{(-\sqrt{6}/4)(a-x)}^{(\sqrt{6}/4)(a+x)} \int_0^{\sqrt{2}(a-|(u/\sqrt{6})+x|)} e^{-\frac{1}{2}(u^2+t^2)}\, dt\, du, \\ \qquad\qquad \text{for } 0 \leqq |x| \leqq \dfrac{a}{3}, \\ \dfrac{\sqrt{6}c^3}{2\pi^{3/2}} e^{-(3/2)x^2} \displaystyle\int_{(-\sqrt{6}/4)(a-x)}^{\sqrt{6}(a-x)} \int_0^{\sqrt{2}(a-|(u/\sqrt{6})+x|)} e^{-\frac{1}{2}(u^2+t^2)}\, dt\, du, \\ \qquad\qquad \text{for } \dfrac{a}{3} < |x| < a, \\ 0 \quad \text{otherwise.} \end{cases} \tag{3.3}$$

Case $n = 4$. Applying the convolution law to the density (3.2) it is found that

$$f_4(x) = \begin{cases} \dfrac{4}{\pi^2} c^4 e^{-2x^2} \displaystyle\int_0^{2(a-|x|)} \int_0^{\sqrt{2}(a-|(u/2)+x|)} \int_0^{\sqrt{2}(a-|(u/2)-x|)} \\ \qquad \cdot e^{-\frac{1}{2}(u^2+v^2+w^2)}\, dw\, dv\, du, \quad \text{for } |x| < a, \\ 0 \quad \text{otherwise.} \end{cases} \tag{3.4}$$

For sufficiently large n, Birnbaum and Andrews [1] have pointed out that $n\bar{X}$ has a limiting normal distribution. Thus for large n one may obtain an approximate cumulative distribution of $\bar{X}$ from (4.2) in [1]. However, in this paper we shall confine our attention to only those cases where $n \leqq 4$.

4. Tests of hypotheses under truncation. In this section we consider the effect of truncation on size and power of tests of hypotheses concerning the means of parent populations.

TABLE I

Values of $P_u(\mu)$, $P(\mu, a)$, $L(\mu, a)$ and loss in power expressed as percentage of $P_u(\mu)$ for $\alpha = 0.05$

n	a	1.5						2.0					
	μ	0	0.5	1.0	1.5	2.0	2.5	0	1.0	1.5	2.0	2.5	3.0
1	P_u	.0500	.1261	.2595	.4424	.6387	.8037	.0500	.2595	.4424	.6387	.8037	.9123
	P	.0000	.0684	.2224	.4335	.6601	.8506	.0285	.2480	.4396	.6453	.8182	.9319
	L	.0500	.0577	.0371	.0089	−.0214	−.0469	.0215	.0115	.0028	−.0066	−.0145	−.0196
	% Loss	100.0	45.8	14.3	2.0	−3.4	−5.8	43.0	4.4	0.6	−1.0	−1.8	−2.1
2	P_u	.0500	.1742	.4087	.6831	.8817	.9707	.0500	.4087	.6831	.8817	.9707	.9953
	P	.0094	.1108	.3839	.7289	.9422	.9984	.0297	.4005	.6989	.9068	.9864	.9997
	L	.0406	.0634	.0248	−.0458	−.0605	−.0277	.0203	.0082	−.0158	−.0251	−.0157	−.0044
	% Loss	81.2	36.4	6.1	−6.7	−6.9	−2.9	40.6	2.0	−2.3	−2.8	−1.6	−0.4
3	P_u	.0500	.2180	.5347	.8297	.9656	.9964	.0500	.5347	.8297	.9656	.9964	.9998
	P	.0112	.1533	.5449	.8961	.9951	1.0	.0301	.5384	.8566	.9820	.9996	1.0
	L	.0388	.0647	−.0102	−.0664	−.0295	−.0036	.0199	−.0037	−.0269	−.0164	−.0032	−.0002
	% Loss	77.6	29.7	−1.9	−8.0	−3.1	−0.4	39.8	−0.7	−3.2	−1.7	−0.3	—
4	P_u	.0500	.2595	.6387	.9123	.9907	.9996	.0500	.6387	.9123	.9907	.9996	1.0−
	P	.0118	.1977	.6792	.9664	.9998	1.0	.0303	.6539	.9374	.9972	1.0−	1.0
	L	.0382	.0618	−.0405	−.0541	−.0091	−.0004	.0197	−.0152	−.0251	−.0065	−.0004	—
	% Loss	76.4	23.8	−6.3	−5.9	−0.9	—	39.4	−2.4	−2.8	−0.7	—	—

n	σ	2.5						3.0					
	μ	0	0.75	1.50	2.25	3.00	3.75	0	0.5	1.0	2.0	3.0	4.0
1	P_u	.0500	.1854	.4424	.7274	.9123	.9824	.0500	.1261	.2595	.6387	.9123	.9907
	P	.0443	.1814	.4416	.7303	.9175	.9884	.0488	.1251	.2588	.6391	.9134	.9921
	L	.0057	.0040	.0008	−.0029	−.0052	−.0060	.0012	.0010	.0007	−.0004	−.0011	−.0014
	% Loss	11.4	2.2	0.2	−0.4	−0.6	−0.6	2.4	0.8	0.3	−0.1	−0.1	−0.1
2	P_u	.0500	.2795	.6831	.9379	.9953	.9999	.0500	.1742	.4087	.8817	.9953	1.0−
	P	.0426	.2742	.6875	.9456	.9978	1.0	.0480	.1725	.4082	.8826	.9962	1.0−
	L	.0074	.0053	−.0044	−.0077	−.0025	−.0001	.0020	.0017	.0005	−.0009	−.0009	—
	% Loss	14.8	1.9	−0.6	−0.8	−0.3	—	4.0	1.0	0.1	−0.1	—	—
3	P_u	.0500	.3647	.8297	.9878	.9998	1.0−	.0500	.2180	.5347	.9656	.9998	1.0−
	P	.0425	.3605	.8348	.9917	1.0−	1.0−	.0479	.2161	.5350	.9675	.9999	1.0
	L	.0075	.0042	−.0051	−.0039	−.0002	—	.0021	.0019	−.0003	−.0019	−.0001	—
	% Loss	15.0	1.2	−0.6	−0.4	—	—	4.2	0.9	−0.1	−0.2	—	—
4	P_u	.0500	.4424	.9123	.9978	1.0−	1.0−	.0500	.2595	.6387	.9907	1.0−	1.0−
	P	.0425	.4403	.9213	.9980	1.0−	1.0	.0478	.2575	.6399	.9918	1.0−	1.0
	L	.0075	.0021	−.0090	−.0011	—	—	.0022	.0020	−.0012	−.0011	—	—
	% Loss	15.0	0.5	−1.0	−0.1	—	—	4.4	0.8	−0.2	−0.1	—	—

Consider a sample of size n from a normal distribution $N(\mu, 1)$. Then a Uniformly Most Powerful (UMP) test of the one-sided hypothesis testing problem

$$H: \mu = \mu_0, \qquad \text{Alt}: \mu > \mu_0, \tag{4.1}$$

is given by (we assume without loss of generality that $\mu_0 = 0$),

$$\text{Reject } H \text{ if } \bar{X} > Z_\alpha/\sqrt{n}; \qquad \text{accept } H \text{ otherwise}, \tag{4.2}$$

where Z_α is the point exceeded with probability α using the distribution of the standard normal variable. Now, if sampling from $N_a(\mu, 1)$ where $N_a(\mu, 1)$ is the density (2.1) with $\sigma = 1$, and test procedure (4.2) is used, the predetermined size α of this 'usual' test is really not obtained. The actual size is given by $\alpha' = \Pr(Z_t > Z_\alpha/\sqrt{n})$, where Z_t is the random variable with density function $f_a(x)$ of the last section.

Now, the 'usual' power function of the test (4.2) is given by

$$\begin{aligned} P_u(\mu) &= \Pr\{\bar{X} > Z_\alpha/\sqrt{n} \mid \bar{X} \sim N(\mu, 1/n)\} \\ &= \Pr\{Z > Z_\alpha - \mu\sqrt{n} \mid Z \sim N(0, 1)\}, \end{aligned} \tag{4.3}$$

if sampling is from a "complete" normal distribution. However, if the sampling is from a truncated distribution, $N_a(\mu, 1)$, the actual power function of the 'usual' size α test is given by

$$\begin{aligned} P(\mu, a) &= \Pr\{\bar{X} > Z_\alpha/\sqrt{n} \mid \bar{X} \sim f_a(x, \mu)\} \\ &= \Pr\{Z_t > Z_\alpha/\sqrt{n} - \mu \mid Z_t \sim f_a(x)\}, \end{aligned} \tag{4.4}$$

where $f_a(x, \mu)$ is the density of $\bar{X}$ when sampling from $N_a(\mu, 1)$, and Z_t is the random variable with density $f_a(x) = f_a(x, 0)$.

We denote the difference of (4.4) and (4.3) by

$$L(\mu, a) = P_u(\mu) - P(\mu, a). \tag{4.5}$$

For $\mu = 0$, L equals $\alpha - \alpha'$, while for all other values of μ, L is the "loss of power" if the usual procedure is followed, while sampling is actually from a truncated distribution. Values of $P_u(\mu)$, $P(\mu, a)$, and the loss in power expressed as percentage of $P_u(\mu)$ for different values of μ and four terminus points 'a' are given in Table I for $\alpha = 0.05$ and $n = 1, 2, 3$ and 4.

It can be easily verified that (4.5) reduces to

$$\begin{aligned} L(\mu, a) = &\Pr\left\{\bar{X} - \mu > \frac{Z_\alpha}{\sqrt{n}} - \mu \,\middle|\, \bar{X} \sim N\left(0, \frac{1}{n}\right)\right\} \\ &- \Pr\left\{\bar{X} - \mu > \frac{Z_\alpha}{\sqrt{n}} - \mu \,\middle|\, \bar{X} \sim f_a(x, 0)\right\}, \end{aligned} \tag{4.6}$$

and by graphical considerations one may see that $L(\mu, a)$ and $Z_\alpha/\sqrt{n} - \mu$ have the same sign. Thus, as soon as μ exceeds $Z_\alpha/\sqrt{n}$, there will be a change of sign from positive to negative in the loss of power, $L(\mu, a)$. This is borne out by the actual computations in Table I.

TABLE II

Upper 100 α% points of $f_n(x)$

α	a	n			
		1	2	3	4
0.10	1.0	.749	.510	.408	.351
	1.5	1.022	.693	.559	.482
	2.0	1.184	.813	.659	.569
	2.5	1.254	.875	.711	.615
	3.0	1.275	.898	.732	.634
0.05	1.0	.868	.636	.516	.445
	1.5	1.226	.871	.708	.613
	2.0	1.472	1.028	.838	.725
	2.5	1.593	1.114	.909	.786
	3.0	1.633	1.150	.938	.812
0.025	1.0	.932	.731	.603	.524
	1.5	1.350	1.011	.831	.722
	2.0	1.679	1.204	.987	.857
	2.5	1.868	1.315	1.076	.933
	3.0	1.939	1.365	1.115	.966
0.01	1.0	.972	.821	.695	.609
	1.5	1.436	1.155	.965	.843
	2.0	1.848	1.396	1.154	1.006
	2.5	2.142	1.545	1.266	1.100
	3.0	2.279	1.611	1.318	1.143
0.005	1.0	.986	.870	.751	.664
	1.5	1.467	1.238	1.049	.922
	2.0	1.919	1.515	1.262	1.104
	2.5	2.285	1.687	1.392	1.212
	3.0	2.493	1.775	1.455	1.263

Now, suppose the sampling is from $N_a(\mu, 1)$. By applying the Neyman-Pearson Fundamental Lemma, a UMP test of (4.1) of size α is

$$\text{(4.7)} \qquad \begin{cases} \text{Reject } H \text{ if } \bar{X} > K_a(a, n), \\ \text{Accept } H \text{ otherwise,} \end{cases}$$

where $K_a(a, n)$ is the point exceeded with probability α using the distribution whose density is $f_n(x)$. Table II gives the significance points for the test (4.7) for different n, α and a. That is, if sampling from a truncated normal distribution, (4.7) gives the 'correct' test for problem (4.1), and Table II gives the correct significance points for this problem.

The power of this 'correct' test (4.7) is given by

$$\text{(4.8)} \qquad P_c(\mu) = \Pr\,(Z_t > K_a(a, n) - \mu),$$

where Z_t is the random variable with density $f_n(x)$. The gain in power, $G(\mu, a) = P_c(\mu) - P(\mu, a)$, is the gain that would result if one uses the correct test rather than the usual test. The values of $P_c(\mu)$, $G(\mu, a)$ and the gain in power expressed

TABLE III

Values of $P_c(\mu)$; $G(\mu, a)$ and gain in power expressed as percentage of $P(\mu, a)$ for $\alpha = 0.05$

n	a	1.5					2.0				
	μ	.5	1.0	1.5	2.0	2.5	1.0	1.5	2.0	2.5	3.0
1	P_c	.1929	.3968	.6246	.8238	.9601	.3098	.5117	.7108	.8646	.9576
	G	.1245	.1744	.1911	.1637	.1095	.0618	.0721	.0655	.0464	.0257
	% Gain	182.0	78.4	44.1	24.8	12.9	24.9	16.4	10.2	5.7	2.8
2	P_c	.2508	.5923	.8761	.9880	1.0	.4828	.7684	.9393	.9935	.9999
	G	.1400	.2084	.1472	.0458	.0016	.0823	.0695	.0325	.0071	.0002
	% Gain	126.4	54.3	20.2	4.9	0.2	20.5	9.9	3.6	0.7	—
3	P_c	.3199	.7450	.9684	.9997	1.0	.6220	.9010	.9905	.9999	1.0
	G	.1666	.2001	.0723	.0046	—	.0836	.0444	.0085	.0003	—
	% Gain	108.7	36.7	8.1	0.5	—	15.5	5.2	0.9	—	—
4	P_c	.3836	.8468	.9931	1.0−	1.0	.7305	.9611	.9988	1.0−	1.0
	G	.1859	.1676	.0267	.0002	—	.0766	.0237	.0016	—	—
	% Gain	94.0	24.7	2.8	—	—	11.7	2.5	0.2	—	—

n	a	2.5					3.0				
	μ	.75	1.50	2.25	3.00	3.75	.5	1.0	2.0	3.0	4.0
1	P_c	.1958	.4625	.7475	.9256	.9906	.1276	.2627	.6436	.9152	.9924
	G	.0144	.0209	.0172	.0081	.0022	.0025	.0039	.0045	.0018	.0003
	% Gain	7.9	4.7	2.4	0.9	0.2	2.0	1.5	0.7	0.2	—
2	P_c	.2986	.7124	.9534	.9983	1.0	.1773	.4156	.8873	.9965	1.0
	G	.0244	.0249	.0078	.0005	—	.0048	.0074	.0047	.0003	—
	% Gain	8.9	3.6	0.8	—	—	2.8	1.8	0.5	—	—
3	P_c	.3881	.8558	.9933	.9999	1.0	.2221	.5431	.9689	.9999	1.0
	G	.0276	.0210	.0016	—	—	.0060	.0081	.0014	—	—
	% Gain	7.7	2.5	0.2	—	—	2.8	1.5	0.1	—	—
4	P_c	.4703	.9320	.9992	1.0−	1.0	.2644	.6477	.9922	1.0−	1.0
	G	.0300	.0107	.0003	—	—	.0069	.0078	.0004	—	—
	% Gain	6.8	1.2	—	—	—	2.7	1.2	—	—	—

as percentage of $P(\mu, a)$ for different μ, a and n and for $\alpha = 0.05$ are given in Table III.

5. Conclusions. An examination of the tables indicates that serious losses occur in the size of the test rather than its power. For example, if the truncation occurs at 1.5 times the standard deviation on either side of the mean, and a usual 5% significance test is used, one is really using approximately 1% significance test rather than 5%. If the truncation occurs at twice the standard deviation on

either side of the mean, the usual 5% significance test gives only approximately 3% significance level. Thus one consequence of applying the usual test is to err on the conservative side in making it much more difficult to reject the hypothesis. As expected, however, when the truncation is at about three times the standard deviation on either side of the mean, there is hardly any difference between the usual and the correct test. Even when the truncation occurs at less than twice the standard deviation away from the mean, there is not much change in the value of the power function beyond one standard deviation away from the value of the mean specified by the null hypothesis. Hence it would appear that unless there is severe truncation and unless the alternative value of the mean is quite near the value specified by the null hypothesis, the usual test would be satisfactory. The results given here are only for a usual 5% significance level test. It is proposed to give extensive tables of the distribution of the mean of samples from truncated distributions and to examine the tests at other than 5% significance levels in another paper.

REFERENCES

[1] Z. W. Birnbaum, and F. C. Andrews, "On sums of symmetrically truncated normal random variables," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 458–461.

[2] Francis L. Campbell, "A study of truncated bivariate normal distributions," Doctoral Dissertation, University of Michigan, (June, 1945).

[3] Douglas G. Chapman, "Estimating the parameters of a truncated gamma distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 498–506.

[4] A. C. Cohen, Jr., "On estimating the mean and standard deviation of truncated normal distributions," *J. Amer. Stat. Assn.*, Vol. 44 (1949), pp. 518–525.

[5] A. C. Cohen, Jr., "Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 557–569.

[6] A. C. Cohen, Jr., "Estimating parameters of Pearson type III populations from truncated samples," *J. Amer. Stat. Assn.*, Vol. 45 (1950), pp. 411–423.

[7] A. C. Cohen, Jr., "Estimation of parameters in truncated Pearson frequency distributions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 256–265.

[8] A. C. Cohen, Jr., "Estimation in truncated bivariate normal distributions," *University of Georgia, Mathematical Technical Report No. 2*, Contract DA-01-009-ORD-288, (June, 1953).

[9] A. C. Cohen, Jr., "Estimation in truncated multivariate normal distributions," *University of Georgia, Mathematical Technical Report No. 3*, Contract DA-01-009-ORD-288 (August, 1953).

[10] A. C. Cohen, Jr., "Restriction and selection in samples from bivariate normal distributions," *J. Amer. Stat. Assn.*, Vol. 50 (1955), pp. 884–893.

[11] A. Clifford Cohen, Jr., "Restriction and selection in multinormal distributions," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 731–741.

[12] A. C. Cohen, Jr., and John Woodward, "Tables of Pearson-Lee-Fisher Functions of singly truncated normal distributions," *Biometrics*, Vol. 9 (1953), pp. 489–497.

[13] Walter L. Deemer, and David F. Votaw, Jr., "Estimation of parameters of truncated or censored experimental distributions," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 498–504.

[14] George Gerard den Broeder, "On parameter estimation for truncated Pearson type III distributions," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 659–663.

[15] D. J. Finney, "The truncated binomial distribution." *Ann. Eugenics*, Vol. 14 (1949), pp. 319–328.

[16] V. J. Francis, "On the distribution of the sum of n sample values drawn from a truncated normal population," *J. Roy. Stat. Soc. Suppl.*, Vol. 8 (1946), pp. 223–232.

[17] A. Hald, "Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point," *Skand. Aktuarietids*, Vol. 32 (1949), pp. 119–134.

[18] Harold Hotelling, "Fitting generalized truncated normal distributions," Abstracts of Madison meeting, *Ann. Math. Stat.*, Vol. 19 (1948), p. 596.

[19] Karl Pearson, "On the influence of natural selection on the variability and correlation of organs," *Philos. Trans. Roy. Soc. London Ser. A*, Vol. 200 (1903), pp. 1–66.

[20] Des Raj, "Estimation of the parameters of type III populations from truncated samples," *J. Amer. Stat. Assn.*, Vol. 48 (1953), pp. 336–349.

[21] Des Raj, "On estimating the parameters of bivariate normal populations from doubly and singly, linearly truncated samples," *Sankhya*, Vol. 12 (1953), pp. 277–290.

[22] Walter L. Smith, "A note on truncation and sufficient statistics," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 247–252.

[23] John W. Tukey, "Sufficiency, truncation and selection," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 309–311.

[24] John W. Tukey, "The Truncated Mean in Moderately Large Samples," *Memorandum Report 32*, Statistical Research Group, Princeton University (1949).

[25] D. F. Votaw, Jr., J. A. Rafferty, and W. L. Deemer, "Estimation of parameters in a truncated trivariate normal distribution," *Psychometrika*, Vol. 15 (1950), pp. 339–347.

NOTES

BARTLETT DECOMPOSITION AND WISHART DISTRIBUTION

By A. M. Kshirsagar

University of Bombay

1. Introduction. In a recent paper Wijsman [1] has presented a method of deriving the Bartlett decomposition and the Wishart distribution by using orthogonal matrices depending upon certain random vectors. The method used in the paper was simple compared to other methods in literature. The present paper gives a similar method, depending on orthogonalization of vectors, and is simpler and direct. This method explicitly gives the χ^2 variables and normal variables in the decomposition of the Wishart distribution in a straightforward way.

The device of writing the Wishart matrix as a product of a triangular matrix and its transpose has been used before; see [4] and [6]. In [5], this is done and it is shown that the elements of this matrix are independent chi and normal variables. However, the present method seems simpler in that it leads to the variables via transformations rather than via densities and Jacobians.

2. Notation and results. The same notation as in Wijsman's [1] paper is used. Let x_{it} $(i = 1, \cdots, k; t = 1, \cdots, n)$ be independent $N(0, 1)$ variables, forming the $k \times n$ matrix M^2_{kn}, the ith row of which is denoted by X'_i $(i = 1, \cdots, k)$, and let

$$A^2_{kn} = [a_{ij}] = M^2_{kn}(M^2_{kn})'. \tag{1}$$

Obviously $a_{ij} = X'_i X_j$ $(i, j = 1, \cdots, k)$. Orthogonalizing the vectors $X_1, \cdots, X_k$, we get vectors

$$Y_i = X_i - b'_{i1}Y_1 - b'_{i2}Y_2 - \cdots - b'_{i,i-1}Y_{i-1} \quad i = 1, \cdots, k \tag{2}$$

where b'_{ir} $(i = 2, \cdots, k; r = 1, \cdots, i-1)$ are so chosen that

$$Y'_i Y_j = 0, \quad i \neq j; \quad i, j = 1, \cdots, k. \tag{3}$$

It is therefore easy to see that $b'_{ir} = Y'_r X_i / Y'_r Y_r$. Let

$$b_{ir} = \frac{Y'_r X_i}{(Y'_r Y_r)^{1/2}} = (Y'_r Y_r)^{1/2} b'_{ir}. \qquad \begin{matrix} i = 2, \cdots, k \\ r = 1, \cdots, i-1. \end{matrix} \tag{4}$$

On account of (3), it can be seen that

$$X'_i X_i = Y'_i Y_i + \sum_{r=1}^{i-1} b^2_{ir} \tag{5}$$

and

$$X'_i X_j = \sum_{r=1}^{i-1} b_{ir} b_{jr} + b_{ji}(Y'_i Y_i)^{1/2} \quad \text{for} \quad j > i \qquad i, j = 1, \cdots, k.$$

Received February 4, 1958.

This can be written, in matrix form as

$$A_{in}^{s} = B_i B_i' \qquad i = 1, \cdots, k, \tag{6}$$

where B_i is the triangular matrix

$$\begin{bmatrix} (Y_1'Y_1)^{1/2} & 0 & 0 & \cdots & 0 \\ b_{21} & (Y_2'Y_2)^{1/2} & 0 & \cdots & 0 \\ b_{31} & b_{32} & (Y_3'Y_3)^{1/2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{i1} & b_{i2} & b_{i3} & \cdots & (Y_i'Y_i)^{1/2} \end{bmatrix}. \tag{7}$$

Taking determinants on both sides of (6), we have

$$|A_{in}^{s}| = (Y_1'Y_1)(Y_2'Y_2)\cdots(Y_i'Y_i) \qquad i = 1, \cdots, k \tag{8}$$

and

$$Y_i'Y_i = \frac{|A_{in}^{s}|}{|A_{i-1n}^{s}|}, \qquad i = 1, \cdots, k \tag{9}$$

where by convention $A_{0n}^{s} = 1$.

Consider the conditional distribution of b_{ir} $(r = 1, \cdots, i-1)$ when $X_1, \cdots, X_{i-1}$ are fixed. The b_{ir} are $i-1$ orthogonal (on account of (3)) linear functions of the independent $N(0, 1)$ variables $x_{i1}, x_{i2}, \cdots, x_{in}$ and

$$E(b_{ir}) = 0$$

$$V(b_{ir}) = \frac{Y_r'Y_r}{Y_r'Y_r} = 1.$$

Therefore the b_{ir} $(r = 1, \cdots, i-1)$ are independent $N(0, 1)$ and by Fisher's lemma (See [2]),

$$Y_i'Y_i = X_i'X_i - \sum_{r=1}^{i-1} b_{ir}^2$$

is a χ^2 variable with $n-(i-1)$ degrees of freedom. This is independent of the normal distribution of b_{ir}. Since the fixed variables $X_1, \cdots, X_{i-1}$ do not appear in the distribution of $Y_i'Y_i$ and b_{ir} $(r = 1, \cdots, i-1)$, and since the result is true for every i, we get the following result:

$Y_i'Y_i$ $(i = 1, \cdots k)$ are independent χ^2 variables with $n-(i-1)$ degrees of freedom, and b_{ir} $(i = 2, \cdots, k;\ r = 1, \cdots, i-1)$ are independent $N(0, 1)$ variables, and all these variables are independently distributed.

The density factor in the joint distribution of the $\frac{1}{2}k(k+1)$ variables $Y_i'Y_i$ and b_{ir} is therefore

$$C_{kn} \exp\left[-\frac{1}{2}\sum_{i=2}^{k}\sum_{r=1}^{i-1} b_{ir}^2\right] \prod_{i=1}^{k} \{(Y_i'Y_i)^{(n-i-1)/2} \exp[-\tfrac{1}{2}Y_i'Y_i]\} \tag{10}$$

where

$$C_{kn} = 2^{nk/2}\pi^{k(k-1)/4}\prod_{i=1}^{k}\Gamma\left(\frac{n-i+1}{2}\right). \tag{11}$$

This density factor, can be written, on account of (5) and (8), as

$$C_{kn}\exp\left[-\tfrac{1}{2}\operatorname{tr} A^{x}_{kn}\right]|A^{x}_{kn}|^{(n-k-1)/2}\prod_{i=1}^{k}(Y_i'Y_i)^{(k-i)/2}. \tag{12}$$

If we transform from the variables $Y_i'Y_i$ $(i = 1, \cdots, k)$ and b_{ir} $(i = 2, \cdots, k;$ $r = 1, \cdots, i-1)$ to the $k(k+1)/2$ distinct elements a_{ij} of $A^{x}_{k,n}$, by using the transformation as given by (6), the Jacobian of transformation is (see [3] Theorem 4.1)

$$\prod_{i=1}^{k}(Y_i'Y_i)^{(i-k)/2}$$

Consequently the density factor in the distribution of the elements of $A^{x}_{k,n}$ is

$$C_{kn}|A^{x}_{kn}|^{(n-k-1)/2}\exp\left[-\tfrac{1}{2}\operatorname{tr} A^{x}_{kn}\right]. \tag{13}$$

REMARKS. The $\frac{1}{2}k(k-1)$ normal variables, k χ^2 variables and the triangular matrix in the Bartlett decomposition, as mentioned by Wijsman [1], are directly given by (6).

Further, from (8), (9) and the distribution of $Y_i'Y_i$ as derived above, it follows immediately that

(i) $\dfrac{|A^{x}_{in}|}{|A^{x}_{i-1n}|}$ are independent χ^2 variables

with $n-(i-1)$ degrees of freedom $(i = 1, \cdots, k)$. From this, the results of Lemma 2 and 3 in Wijsman's [1] paper follow very easily.

The author is very grateful to the referee for his valuable suggestions.

REFERENCES

[1] R. A. WIJSMAN, "Random orthogonal transformations," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 415–422.

[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton 1946, page 379.

[3] W. L. DEEMER AND I. OLKIN, "Jacobians of certain matrix transformations used in multivariate analysis," *Biometrika*, Vol. 38 (1951), pp. 345–367.

[4] P. C. MAHALANOBIS, R. C. BOSE, AND S. N. ROY, "Normalization of statistical variates and the use of rectangular co-ordinates in the theory of sampling distributions," *Sankhyā*, Vol. 3 (1937), pp. 1–40.

[5] J. G. MAULDON, "Pivotal quantities for Wishart's and related distributions and a paradox in fiducial theory," *J. Roy. Stat. Soc.*, Ser. B, Vol. 17 (1955), pp. 79–85.

[6] I. OLKIN, "On distribution problems in multivariate analysis," *Inst. of Stat. Mimeo Series*, North Carolina, Vol. 43 (1951), pp. 1–126.

SOME BOUNDS ON THE DISTRIBUTION FUNCTIONS OF THE LARGEST AND SMALLEST ROOTS OF NORMAL DETERMINANTAL EQUATIONS[1]

RAY MICKEY

Iowa State College[2]

While the joint density function of the roots of certain determinantal equations have been obtained, [1], [2], [3], the result is sufficiently complex that the marginal distribution functions of these statistics have not, to the author's knowledge, been tabulated. We present here a lower bound on the distribution function of the smallest root and an upper bound on the distribution function of the largest root.

These bounds may be of possible usefulness in problems of significance tests since observed values that are not "significant" according to the bounds will certainly not be "significant" with respect to the exact distribution.

Let S^1_{ij} and S^2_{ij}, $i, j = 1, \cdots, k$, be two sample covariance matrices from normal distributions having identical covariance matrices. It is well known [4] that the smallest and largest roots, say W_1 and W_k, of the equation

$$|S^1_{ij} - WS^2_{ij}| = 0$$

satisfy the inequalities

$$W_1 \leqq \frac{\sum S^1_{ij} x_i x_j}{\sum S^2_{ij} x_i x_j} \leqq W_k, \qquad \sum S^2_{ij} x_i x_j > 0$$

Let $F_i = S^1_{ii} / S^2_{ii}$. It then follows that

$$W_1 \leqq \min_i \{F_i\}; \qquad W_k \geqq \max_i \{F_i\}.$$

Since the roots are invariant under linear transformations of the underlying variab'es, the covariance matrix may be taken to be the identity matrix. Then the F_i are independently and identically distributed according to the well known F distribution. Denote by $F_{[1]}$ and $F_{[k]}$ the smallest and the largest of a set of k independently identically distributed F values. We then have the desired bounds.

$$P\{W_1 \leqq u\} \geqq P\{F_{[1]} \leqq u\}$$

$$P\{W_k \geqq v\} \geqq P\{F_{[k]} \geqq v\}.$$

Denote by $G(F)$ the distribution function of F (which depends, of course, on the numbers of degrees of freedom for S^1_{ij} and S^2_{ij}). The above bounds become

$$P\{W_1 \leqq u\} \geqq 1 - [1 - G(u)]^k$$

$$P\{W_k \geqq v\} \geqq 1 - [G(v)]^k.$$

Received April 28, 1958.

[1] Prepared in connection with work under AEC contract AT(30-1)-1377.

[2] Now at General Analysis Corporation, Los Angeles, California.

REFERENCES

[1] R. A. Fisher, "The sampling distribution of some statistics obtained from non-linear equations," *Annals of Eugenics*, Vol. 9 (1939), pp. 238–249.

[2] P. L. Hsu, "On the distribution of roots of certain determinantal equations," *Annals of Eugenics*, Vol. 9 (1939), pp. 250–258.

[3] A. M. Mood, "On the distribution of the characteristic roots of normal second-moment matrices," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 266–274.

[4] S. N. Roy, "p-Statistics, or some generalizations on the analysis of variance appropriate to multivariate problems," *Sankhyā*, Vol. 3 (1939), pp. 341–396.

NOTE ON A MOVING SINGLE SERVER PROBLEM[1]

By S. Karlin, R. G. Miller, Jr., and N. U. Prabhu

Stanford University and *Karnatak University*

1. Introduction and summary. B. McMillan and J. Riordan in [1] derived the generating function for the probability distribution of the number of items completed before absorption in a moving single server problem in two special cases. Through an analogy to the work of L. Takács [2] on busy period problems for a simple queue, McMillan and Riordan postulated a nonlinear integral equation relation for the generating function. In this note the validity of this relation is proved in general by exploiting the analogy more fully, and the generating function in the two special cases is obtained directly from the integral equation. A similar functional relation is established for the Laplace-Stieltjes transform of the distribution of time until absorption, and the transform is obtained for the two special cases.

2. Functional relations. As stated by McMillan and Riordan the moving single server problem is the following: an assembly line moving with uniform speed has items for service spaced along it. The single server available moves with the line while serving and against it with infinite velocity while transferring service to the next item in line. The line has a barrier in which the server may be said to be "absorbed" in the sense that service is disabled if the server moves into the barrier. The server with exponentially (α) distributed service time starts service on the first item when it is T time units away from the barrier. Let the spacings between items be independent random variables with the general distribution function $B(t)$.

This problem is analogous to a simple queue with a single server, Poisson arrivals ($\lambda = \alpha$), and distribution of service times $F_S(t) \equiv B(t)$. The time until absorption in the moving single server problem is equivalent to the length of a

Received April 28, 1958.

[1] This paper represents work done independently by Karlin and Miller and by Prabhu, and combined because of the similarity of the results. The work of the first two authors was sponsored in part by Stanford University Office of Naval Research Contract Nonr-225(28).

busy period for the simple queue in which the service distribution for the first item in line is

$$F_T(t) = \begin{cases} 1, & t > T, \\ 0, & t \leqq T. \end{cases}$$

The number of items completed before absorption is one less than the number of items serviced during a busy period in which the first item has the service distribution F_T.

Let $p(k, T)$ be the probability that the server completes k items before absorption, and let $P(x, T) = \sum_{k=0}^{\infty} p(k, T)x^k$. Let f_j be the probability that j items are serviced during a busy period in which the first item has service distribution F_S, and let $F(x) = \sum_{j=1}^{\infty} f_j x^j$.

For the queue suppose that n items arrive during the service period (of length T) of the first item. The probability distribution on the number of items serviced during the remainder of the busy period is the n-fold convolution of $f = \{f_j\}$. Hence,

$$P(x, T) = \sum_{n=0}^{\infty} \frac{e^{-\alpha T}(\alpha T)^n}{n!} [F(x)]^n = e^{-\alpha T(1-F(x))}, \tag{1}$$

where $F(x)$ is defined above and is the unique analytic solution to the integral equation

$$F(x) = x \int_0^{\infty} e^{-\alpha t(1-F(x))}\, dB(t), \qquad |x| \leqq 1, \tag{2}$$

subject to the condition $F(0) = 0$ (see [2]). Since the integrand in (2) is $P(x, t)$, $P(x, t)$ is the unique analytic solution to the non-linear integral equation

$$P(x, T) = \exp\left\{-\alpha T\left(1 - x\int_0^{\infty} P(x, t)\, dB(t)\right)\right\}, \tag{3}$$

subject to the condition $|P(x, t)| \leqq 1$ for $|x| \leqq 1$, all t. This is the integral relation conjectured by McMillan and Riordan.

Let $H(u, T)$ be the probability that the server is absorbed prior to time u, and let $\tilde{H}(s, T) = \int_0^{\infty} e^{-su}\, dH(u, T)$. Let G be the distribution of the length of a busy period in which the initial item has the service distribution F_S, and let $\tilde{G}(s) = \int_0^{\infty} e^{-su}\, dG(u)$.

The length of time until absorption is T plus an n-fold convolution of busy periods where n is the number of items arriving in the interval $(0, T)$. Hence

$$H(u) = \begin{cases} \sum_{n=0}^{\infty} \frac{e^{-\alpha T}(\alpha T)^n}{n!} G^{(n)}(u - T), & u > T, \\ 0, & u \leqq T, \end{cases} \tag{4}$$

where $G^{(n)}$ denotes the n-fold convolution of G. In terms of Laplace-Stieltjes transforms (4) becomes

$$\tilde{H}(s, T) = \exp\{-T(s + \alpha(1 - \tilde{G}(s)))\}, \tag{5}$$

where $\tilde{G}(s)$ is the unique analytic solution to the relation

$$\tilde{G}(s) = \int_0^\infty e^{-(s+\alpha(1-\tilde{G}(s)))t}\, dB(t), \tag{6}$$

subject to the condition $\lim_{s\to+\infty} \tilde{G}(s) = 0$ for real s (see [2]). Combination of (5) and (6) implies that $\tilde{H}(s, t)$ is the unique analytic solution to the equation

$$\tilde{H}(s, T) = \exp\left\{-sT - \alpha T\left(1 - \int_0^\infty \tilde{H}(s, t)\, dB(t)\right)\right\}, \tag{7}$$

subject to the conditions $|\tilde{H}(s, t)| \leqq 1$ for $\mathrm{Re}\{s\} > 0$, all t and $\lim_{s\to\infty} \tilde{H}(s, t) = 0$ for real s, all t.

3. Examples.

(a)
$$B(t) = \begin{cases} 1, & t > \epsilon, \\ 0, & t \leqq \epsilon. \end{cases}$$

$P(x, T)$ can be determined either from (1) and (2) or from (3) directly. To determine $P(x, T)$ from (3) let $T = \epsilon$ in (3).

$$P(x, \epsilon) = e^{-\alpha\epsilon(1-xP(x,\epsilon))} \tag{8}$$

so $P(x, \epsilon)$ satisfies the equation

$$\alpha\epsilon x P(x, \epsilon) e^{-\alpha\epsilon x P(x,\epsilon)} = \alpha\epsilon x e^{-\alpha\epsilon}. \tag{9}$$

The expansion of $e^{sT/\epsilon}$ for $z = \alpha\epsilon x P(x, \epsilon)$ (see [1]) is

$$e^{sT/\epsilon} = 1 + \sum_{k=1}^{\infty} \frac{(T/\epsilon)(T/\epsilon + k)^{k-1}}{k!} (\alpha\epsilon x e^{-\alpha\epsilon})^k \tag{10}$$

so

$$P(x, T) = e^{-\alpha T} + \sum_{k=1}^{\infty} \frac{T(T + k\epsilon)^{k-1}}{k!} e^{-\alpha T} (\alpha e^{-\alpha\epsilon})^k x^k. \tag{11}$$

$\tilde{H}(s, T)$ can be determined from (11), from (5) and (6), or from (7). For the latter method let $T = \epsilon$ in (7).

$$\alpha\epsilon \tilde{H}(s, \epsilon) e^{-\alpha\epsilon\tilde{H}(s,\epsilon)} = \alpha\epsilon e^{-\epsilon(s+\alpha)} \tag{12}$$

so the expansion of $e^{\alpha T\tilde{H}(s,\epsilon)}$ is

$$e^{\alpha T\tilde{H}(s,\epsilon)} = 1 + \sum_{k=1}^{\infty} \frac{(T/\epsilon)(T/\epsilon + k)^{k-1}}{k!} (\alpha\epsilon e^{-\epsilon(s+\alpha)})^k \tag{13}$$

and

$$\bar{H}(s, T) = e^{-(s+\alpha)T} + \sum_{k=1}^{\infty} \frac{T(T + k\epsilon)^{k-1}}{k!} e^{-\alpha T}(\alpha e^{-\alpha\epsilon})^k (e^{-s(T+k\epsilon)}). \tag{14}$$

$$B(t) = 1 - e^{-\beta t}, \qquad t \geqq 0, \beta > 0. \tag{b}$$

To determine $P(x, T)$ from (3) integrate both sides of (3) with respect to $dB(T)$ and solve for $\int_0^\infty P(x, t)\, dB(t)$.

$$\int_0^\infty P(x, t)\, dB(t) = \frac{\alpha + \beta - \sqrt{(\alpha + \beta)^2 - 4\alpha\beta x}}{2\alpha x} \tag{15}$$

so

$$P(x, T) = \exp\left\{-\frac{T}{2}\left(\alpha - \beta + \sqrt{(\alpha + \beta)^2 - 4\alpha\beta x}\right)\right\}. \tag{16}$$

To determine $\bar{H}(s, T)$ from (7) integrate both sides of (7) with respect to $dB(T)$ and solve for $\int_0^\infty \bar{H}(s, t)\, dB(t)$.

$$\int_0^\infty \bar{H}(s, t)\, dB(t) = \frac{s + \alpha + \beta - \sqrt{(s + \alpha + \beta)^2 - 4\alpha\beta}}{2\alpha} \tag{17}$$

so

$$\bar{H}(s, T) = \exp\left\{-\frac{T}{2}\left(s + \alpha - \beta + \sqrt{(s + \alpha + \beta)^2 - 4\alpha\beta}\right)\right\}. \tag{18}$$

REFERENCES

[1] B. McMillan and J. Riordan, "A moving single server problem," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 471–478.

[2] L. Takács, "Investigation of waiting time problems by reduction to Markov processes," *Acta Mathematica, Acad. Scient. Hung.*, Vol. 6 (1955), pp. 101–128.

DISTRIBUTION OF THE "BLOCKS ADJUSTED FOR TREATMENTS" SUM OF SQUARES IN INCOMPLETE BLOCK DESIGNS

By A. M. Kshirsagar

Bombay University

Introduction. Marvin Zelen [1] has stated that the distribution of the "blocks adjusted for treatments" sum of squares in an incomplete block design is unknown. The present paper is intended to derive this distribution.

Notation and derivation. Let there be v treatments and b blocks having $k_1, k_2, \cdots, k_b$ plots respectively and let the ith treatment be replicated r_i times; $(i = 1, 2, \cdots, v)$. Let n_{ij} (which is either zero or one) be the number of times the ith treatment occurs in the jth block, $(i = 1, 2, \cdots v; j = 1, 2, \cdots, b)$. Then

Received May 5, 1958.

$$\underset{v\times b}{N} = [n_{ij}]$$

is the incidence matrix of the design. Let the total yields of the blocks be denoted by $B_1, B_2, \cdots, B_b$ respectively and the total yields of the treatments by $T_1, T_2, \cdots, T_v$ respectively. Let

$$B = \begin{bmatrix} B_1 \\ B_2 \\ \cdots \\ B_b \end{bmatrix}, \qquad T = \begin{bmatrix} T_1 \\ T_2 \\ \cdots \\ T_v \end{bmatrix}.$$

We shall assume that the yield of a plot consists of a general effect μ, the effect of the block containing the plot, the effect of the treatment which is applied to the plot and the error component. Let α_j denote the effect of the jth block, and let

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_b \end{bmatrix}.$$

It is assumed that the errors are independently and normally distributed with zero mean and variance σ^2, the block effects are independently and normally distributed with zero mean and variance σ'^2 and that the block effects and error components are independently distributed of each other.

Let E_1, V_1, cov_1 denote respectively the expectation, variance-covariance matrix and covariance matrix, in the conditional distribution of the yields when the block effects are fixed; and E, V, cov will denote the same quantities, in the absolute distribution of the yields when the block effects are also normally distributed. Let

$$P = B - N' \operatorname{diag}\left(\frac{1}{r_1}, \frac{1}{r_2}, \cdots, \frac{1}{r_v}\right) T,$$

where diag stands for a diagonal matrix, the elements in the diagonal being written in the adjoining bracket. P is the vector of the block totals, adjusted for treatment effects. Let

$$D = \operatorname{diag}(k_1, k_2, \cdots, k_b) - N' \operatorname{diag}\left(\frac{1}{r_1}, \frac{1}{r_2} \cdots, \frac{1}{r_v}\right) N.$$

Then it can be readily seen that

$$E_1(P) = D\alpha, \qquad V_1(P) = \sigma^2 D.$$

Since the sum of all the elements in any row (or column) of D is zero, one of its latent roots is zero. If the design is a connected one (i.e., all the treatment

contrasts and block contrasts are estimable), the rank of D is $b-1$. Let the non-zero latent roots of D be $\lambda_1, \lambda_2, \cdots, \lambda_{b-1}$, the corresponding orthogonal normalized latent vectors (column vectors) being $m_1, m_2, \cdots, m_{b-1}$. Then $(D-\lambda_i I)m_i = 0$, $i = 1, 2, \cdots, b-1$; and hence

$$\begin{aligned} E_1(m_i'P) &= m_i'D\alpha = \lambda_i m_i'\alpha, && i = 1, 2, \cdots, b-1; \\ V_1(m_i'P) &= \sigma^2 m_i'Dm_i = \sigma^2\lambda_i, && i = 1, 2, \cdots, b-1; \\ \operatorname{cov}_1(m_i'P, m_j'P) &= \sigma^2 m_i'Dm_j && (i \neq j) \\ &= 0 && i, j = 1, 2, \cdots, b-1. \end{aligned}$$

Hence the "blocks adjusted for treatments," sum of squares with $b-1$ degrees of freedom is

$$u = \sum_{i=1}^{b-1} (m_i'P)^2/\lambda_i .$$

When the block effects α_j are not fixed but are random variables obeying a normal distribution with zero mean and variance σ'^2, the means, variances and covariances of $m_i'P$ can be found as below by using theorems 14 and 15 about expectations, variances and covariances proved in [2].

$$\begin{aligned} V(m_i'P) &= (EV_1 + VE_1)m_i'P \\ &= E(\sigma^2\lambda_i) + V(\lambda_i m_i'\alpha) \\ &= \sigma^2\lambda_i + \sigma'^2\lambda_i^2, && i = 1, 2, \cdots b-1; \\ \operatorname{cov}(m_i'P, m_j'P) &= \operatorname{cov}\{E_1(m_i'P), E_1(m_j'P)\} + E\{\operatorname{cov}_1(m_i'P), (m_j'P)\} \\ &= \operatorname{cov}(\lambda_i m_i'\alpha, \lambda_j m_j'\alpha) && (i \neq j) \\ &= 0, && i, j = 1, 2, \cdots b-1; \end{aligned}$$

and

$$\begin{aligned} E(m_i'P) &= E\{E_1(m_i'P)\} \\ &= E(\lambda_i m_i'\alpha) \\ &= 0, && i = 1, 2, \cdots, b-1. \end{aligned}$$

Therefore, if

$$x_i = \frac{m_i'P}{(\sigma^2\lambda_i + \sigma'^2\lambda_i^2)^{1/2}},$$

then $x_1, x_2, \cdots, x_{b-1}$ are standard normal independent variables and the (adjusted) block sum of squares is

$$u = \sum_{i=1}^{b-1} \frac{a_i}{2} x_i^2 ,$$

where

$$a_i = 2(\sigma^2 + \lambda_i\sigma'^2)$$

The distribution of such a quadratic form has been derived by Herbert Robbins [3], and Herbert Robbins and E. J. G. Pitman [4] (Theorem 1).

For a design, the dual of which is a balanced incomplete block design, all the non-zero latent roots of D are equal and the distribution of u reduces to a chi-square distribution.

When the design is not a connected one, some of the λ's will be zero and the necessary changes can be easily made to suit that situation.

I am indebted to Prof. M. C. Chakrabarti, Mr. B. V. Shah of the Bombay University and the referee for valuable suggestions in the preparation of this paper.

REFERENCES

[1] Marvin Zelen, "Analysis of incomplete block designs," *J. Amer. Stat. Assn.*, Vol. 52 (1957), p. 204.

[2] M. H. Hansen, W. N. Hurwitz, and W. G. Madow, *Sample Survey Methods and Theory*, Vol. 2, John Wiley and Sons, New York, pp. 63–65.

[3] Herbert Robbins, "The distribution of a definite quadratic form," *Ann. Math. Stat.*, Vol. 19 (1948), p. 266.

[4] Herbert Robbins, and E. J. G. Pitman, "Application of the method of mixtures to quadratic forms in normal variables," *Ann. Math. Stat.*, Vol. 20 (1949), p. 552.

A SERIES OF SYMMETRICAL GROUP DIVISIBLE INCOMPLETE BLOCK DESIGNS

By D. A. Sprott

Waterloo College, Ontario, Canada

Introduction. A balanced incomplete block design (BIBD) is an arrangement of v elements in b blocks of k different elements each, so that each element occurs in r blocks and each pair of elements occurs in λ blocks. If $v = b(r = k)$, the design is said to be symmetric; it is well known that any two blocks of a symmetric BIBD have exactly λ elements in common.

It has been shown [1] that the subspaces of dimension t of PG(m, p^n) form a BIBD. It is also known that the PG(m, p^n) contains PG(m, p^k) if k is a factor of n (see [2]). In particular, the lines of PG(2, s^2) form the design $v = b = s^4 + s^2 + 1$, $r = k = s^2 + 1$, $\lambda = 1$, which therefore contains the design $v = b = s^2 + s + 1$, $r = k = s + 1$, $\lambda = 1$, (the lines of PG(2, s)), where $s = p^n$.

Received March 18, 1957; revised November 6, 1958.

A group divisible incomplete block design (GD design) has been defined [3] as an arrangement of v elements in b blocks of k different elements, in which the elements can be divided into m groups of n elements each, so that two elements belonging to the same group occur together in λ_1 blocks and two elements belonging to different groups occur together in λ_2 blocks. It will be shown that a series of GD designs can be obtained from the preceding series of BIBD's.

Series of symmetrical group divisible designs. Consider the BIBD (a): $v_1 = b_1 = s^4 + s^2 + 1$, $r_1 = k_1 = s^2 + 1$, $\lambda = 1$. Let set I be the set of blocks of (a) that contains the blocks of the BIBD (b): $v_2 = b_2 = s^2 + s + 1$, $r_2 = k_2 = s + 1$, $\lambda = 1$. Let set II be the remaining blocks of design (a); let set III be the set of blocks remaining in design (a) when all blocks and elements of design (b) have been deleted. Thus set I contains $b_2 = s^2 + s + 1$ blocks, set II contains $b_1 - b_2 = s^4 - s$ blocks, and set III contains $b = v = s^4 - s$ blocks and elements.

LEMMA 1. *Every element of set* III *occurs in exactly* s^2 *blocks of set* III.

PROOF. Since (b) is a symmetric BIBD, any two blocks of set I have exactly one of the $v_2 = s^2 + s + 1$ elements in common. Thus the remaining $(k_1 - k_2)$ $b_2 = s^4 - s$ elements of set I are all different, and they are also different from the v_2 elements of design (b). Since $s^4 - s = v_1 - v_2 = v$, where v is the number of elements in set III, each such element must occur in exactly one block of set I, and so must occur in exactly $r_1 - 1 = s^2$ blocks of set III.

LEMMA 2. *Every block of set* II *contains exactly one element of design* (b).

PROOF. Every pair of elements of design (b) occurs once in set I, and so cannot occur at all in set II. Therefore all elements of design (b) occur in different blocks of set II, each occurring $r_1 - r_2 = s^2 - s$ times. Thus the number of blocks of set II containing exactly one element of design (b) is $v_2(r_1 - r_2) = s^4 - s$, which is the total number of blocks of II.

Combining these lemmas, it can easily be seen that set III contains $v = s^4 - s$ elements in $b = s^4 - s$ blocks, where each block contains $k = k_1 - 1 = s^2$ elements and each element occurs $r = r_1 - 1 = s^2$ times.

Since any element of set III occurs once in set I with $k_1 - k_2 - 1$ other elements of set III, it will not occur again with any of these elements and will occur once in set III with each of the remaining elements. Thus set III is a regular GD design with parameters,

$$v = b = s^4 - s, \qquad r = k = s^2, \qquad \lambda_1 = 0,$$

$$\lambda_2 = 1, \qquad m = s^2 + s + 1, \qquad n = s^2 - s.$$

Example. The lines of PG(2, 4) form a BIBD that contains the design formed by the lines of PG(2, 2). These designs have parameters

(a): $v = b = 21$, $r = k = 5$, $\lambda = 1$, and

(b): $v = b = 7$, $r = k = 3$, $\lambda = 1$ as shown below.

D	*N*	*P*	*Q*	*L*	*O*	*H*	*A*	*A*	*B*	*B*	*C*	*C*	*F*	*F*	*K*	*K*	*J*	*J*	*G*	*G*
E	*R*	*U*	*T*	*M*	*S*	*I*	*N*	*R*	*H*	*I*	*H*	*I*	*D*	*E*	*D*	*E*	*D*	*E*	*D*	*E*
A	*B*	*F*	*C*	*J*	*K*	*G*	*O*	*S*	*L*	*M*	*M*	*L*	*L*	*M*	*I*	*H*	*H*	*I*	*M*	*K*
B	*F*	*C*	*J*	*K*	*G*	*A*	*P*	*T*	*P*	*Q*	*N*	*O*	*Q*	*O*	*N*	*Q*	*O*	*P*	*P*	*N*
C	*J*	*K*	*G*	*A*	*B*	*F*	*Q*	*U*	*T*	*U*	*S*	*R*	*S*	*T*	*T*	*R*	*U*	*S*	*R*	*U*

The lower left hand group of blocks constitutes the design (b), and the lower right hand group of blocks is the GD design with parameters $v = b = 14$, $r = k = 4$, $\lambda_1 = 0$, $\lambda_2 = 1$, $m = 7$, $n = 2$. The groups are (D, E), (N, R), (P, U), (Q, T), (L, M), $O, S)$, and (H, I). Thus, for example, D occurs zero times with E in the GD design and once with N, R, P, U, Q, T, L, M, O, S and I.

A design with these parameters was obtained in [4] using the method of differences and is listed as number R24 in [5]. For $s = 3$ the resulting design has parameters,

$$v = b = 78,\ r = k = 9,\ \lambda_1 = 0,\ \lambda_2 = 1,\ m = 13,\ n = 6.$$

REFERENCES

[1] R. C. Bose, "On the construction of balanced incomplete block designs," *Ann. Eugenics* Vol. 10 (1939), pp. 352–399.

[2] R. D. Carmichael, *Introduction to the Theory of Groups of Finite Order*, Dover Publications, 1956, pp. 334–336.

[3] R. C. Bose, and W. S. Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[4] R. C. Bose, S. S. Shrikhande, and K. N. Bhattacharya, "On the construction of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 167–195.

[5] "Tables of partially balanced designs with two associate classes," Institute of Statistics, University of North Carolina, reprint series No. 50, p. 184 and p. 192.

ALTERNATIVE PROOF OF A THEOREM OF BIRNBAUM AND PYKE

By Nicolaas H. Kuiper

Landbouwhogeschool, Wageningen, Netherlands

Let $U_1, U_2, \cdots, U_n$ be an ordered sample of a random variable (r.v.) X having a uniform distribution (0, 1). If i^* is the value of $i = 1, 2, \cdots, n$ at which $i/n - U_i$ is maximized and $U^* = U_{i^*}$, then U^* is a r.v. with values (0, 1). The probability that the sample cannot be ordered or that i^* is not uniquely defined is zero, and hence these possibilities are neglected. Theorem 3 [1] states that U^* has a uniform distribution (0, 1). Another proof of this fact was given in [2].

Received June 30, 1958.

In this note an alternative proof is given which entails little computation and is self-contained.

Replace the interval (0, 1) by the reals modulo 1, considered as a circle of circumference 1. Let c be an arbitrary point on the circle. Moving from c in the direction corresponding to increasing values (0, 1), one meets successively the points $U_{k+1}, U_{k+2}, \cdots, U_n, U_1, \cdots, U_k$ where k, so defined, is a r.v. depending on c. Rename these points $U_1^c, U_2^c, \cdots, U_n^c$ respectively. Define $i = i(j)$ by $U_j^c = U_i$. Let u_j^c denote the (arc) distance of U_j^c from c taken in the increasing direction. Therefore,

$$i = k + j; \qquad u_j^c = U_{k+j} - c \qquad \text{for } j = 1, \cdots, n - k$$

$$i = k + j - n; \qquad u_j^c = U_{k+j-n} + 1 - c \qquad \text{for } j = n - k + 1, \cdots, n$$

With the indicated relation between i and j observe that

$$j/n - u_j^c = (i - k)/n - U_i + c = i/n - U_i + c - k/n.$$

For a fixed c and a given sample, c and k are constants and hence $j/n - u_j^c$ attains its maximum at the same point $U^* = U_{i^*}$ as does $i/n - U_i$.

Given a sample $U_1, \cdots, U_n$, the point U^* on the circle of reals mod. 1 is therefore independent of the choice of the initial point c taken instead of 0 on this circle. Since the distribution of X mod. 1 is uniform, that is, is invariant under translations, the distribution of U^* mod. 1 is also invariant under translations. Thus U^* has a uniform distribution on (0, 1). q.e.d.

REFERENCES

[1] Z. W. Birnbaum and Ronald Pyke, "On some distributions related to the statistic D_n^+," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 179–187.

[2] Meyer Dwass, "On several statistics related to empirical distribution functions," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 188–191.

QUASI-RANGES OF SAMPLES FROM AN EXPONENTIAL POPULATION

By Paul R. Rider

Wright Air Development Center

In a study of the use of ranges and quasi-ranges in estimating the standard deviation of a population, Harter [4] has compared the results for samples from a normal population with those for samples from certain other populations, including the exponential. In this note are given the distributions of quasi-ranges from the exponential population and also formulas for the cumulants of these quasi-ranges.

Received June 23, 1958; revised October 17, 1958.

Let $x_1, x_2, \cdots, x_n$ be a sample, and suppose that $x_1 \leqq x_2 \leqq \cdots \leqq x_n$. The quasi-range of order r of this sample is the statistic

$$w_r = x_{n-r} - x_{r+1}, \tag{1}$$

w_0 being the range itself.

The population to be considered is

$$f(x) = e^{-x}, \qquad 0 \leqq x, \tag{2}$$

which has mean and variance each equal to 1. The cumulative distribution function for this population is

$$F(x) = \int_0^x e^{-x}\,dx = 1 - e^{-x},$$

and the probability that r members of the sample are below x_{r+1}, r above x_{n-r}, and the remaining values between x_{r+1} and x_{n-r}, is proportional to

$$(1 - e^{-x_{r+1}})^r (e^{-x_{r+1}} - e^{-x_{n-r}})^{n-2r-2} e^{-rx_{n-r}}\,dx_{r+1}\,dx_{n-r}.$$

Replacing x_{n-r} by $x_{r+1} + w_r$, integrating with respect to x_{r+1} between the limits 0 and ∞, and supplying the proper multiplicative constant, we find the distribution of the rth quasi-range, w_r, to be

$$\frac{\Gamma(n-r)}{\Gamma(n-2r-1)\Gamma(r+1)} (1 - e^{-w_r})^{n-2r-2} e^{-(r+1)w_r}\,dw_r. \tag{3}$$

Obviously we must have $n \geqq 2r + 2$.

Upon multiplying (3) by e^{tx} and integrating between the limits 0 and ∞, we have ([2], p. 144) for the moment generating function of the rth quasi-range,

$$M(t) = \frac{\Gamma(n-r)\Gamma(r+1-t)}{\Gamma(r+1)\Gamma(n-r-t)} = \prod_{j=r+1}^{n-r-1} \left(1 - \frac{t}{j}\right)^{-1}. \tag{4}$$

To find the cumulants of the distribution (3) we note that

$$K(t) = \ln M(t) = -\sum_{j=r+1}^{n-r-1} \ln\left(1 - \frac{t}{j}\right) = \sum_{j=r+1}^{n-r-1} \sum_{k=1}^{\infty} \frac{1}{k} \left(\frac{t}{j}\right)^k. \tag{5}$$

It follows at once that the cumulant of order p is

$$\kappa_p = (p-1)! \sum_{j=r+1}^{n-r-1} \frac{1}{j^p}. \tag{6}$$

In particular, we have for the mean and variance respectively,

$$\kappa_1 = \sum_{j=r+1}^{n-r-1} \frac{1}{j}, \qquad \kappa_2 = \sum_{j=r+1}^{n-r-1} \frac{1}{j^2}. \tag{7}$$

Thus the mean of the quasi-range w_r, being equal to the sum of a harmonic series, diverges with sample size, although very slowly, while the variance approaches a finite value. For the case $r = 0$, that is for the range itself, the vari-

ance approaches the value $\pi^2/6 = 1.6449$, and somewhat rapidly. For example, the variance of the range of samples of size 10 is 1.4977.

For $r = 0$, the values of κ_3 and κ_4 approach 2.4041 and $\pi^4/15 = 6.4939$ respectively as n becomes infinite. (Values can be obtained from tables of the Riemann zeta function, e.g. [3].) The ratios $\kappa_3/\kappa_2^{\frac{3}{2}}$ and κ_4/κ_2^2 approach 1.1395 and 12/5 respectively. For a normal distribution these ratios are, of course, both zero.

REFERENCES

[1] J. H. Cadwell, "The distribution of quasi-ranges in samples from a normal population," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 603–613.

[2] Arthur Erdélyi, et al., *Tables of Integral Transforms*, Vol. I, McGraw-Hill Book Co., New York, 1954.

[3] J. P. Gram, *Tafeln für die Riemannsche Zetafunktion*, Høst & Søn, Copenhagen, 1925.

[4] H. Leon Harter, "The Use of Sample Quasi-Ranges in Estimating Population Standard Deviation," *Wright Air Development Center Technical Report* 58–200.

[5] Benjamin Epstein, "Estimates of mean life based on the rth smallest value in a sample of size n drawn from an exponential distribution," *Technical Report No.* 2 (July 1, 1952), prepared under Contract Nonr-451(00) [NR-042-017].

[6] Benjamin Epstein, "Simple estimators of the parameters of exponential distributions when samples are censored," *Ann. Math. Stat.*, Vol. 8 (1956), pp. 15–26.

[7] Shozo Shimada, "Moments of order statistics drawn from an exponential distribution," *Reports of Statistical Application Research*, Union of Japanese Scientists and Engineers, Vol. 4 (1957), pp. 153–158.

ACKNOWLEDGMENT OF PRIORITY

By Ralph G. Stanton

V. N. Murty has kindly pointed out to me that the result of my note, "A Note on Balanced Incomplete Block Designs," (*Ann. Math. Stat.*, Vol. 28 (1957), p. 1054), was given previously by K. Kishen and C. R. Rao in "An Examination of Various Inequality Relations Among Parameters of the Balanced Incomplete Block Design" (*Journal of the Indian Society of Agricultural Statistics*, Vol. IV, No. 2 (1952), pp. 137–144).

CORRECTION TO "RANDOM ORTHOGONAL TRANSFORMATIONS AND THEIR USE IN SOME CLASSICAL DISTRIBUTION PROBLEMS IN MULTIVARIATE ANALYSIS"

By Robert A. Wijsman

In footnote 3 of the paper cited in the title (*Ann. Math. Stat.* Vol. 28 (1957), pp. 415–423), for χ^2 read χ.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Monterey Meeting of the Institute, November 14–15, 1958)

1. On Computing Expectations in Sequential Analysis. FRED C. ANDREWS, University of Oregon, and J. R. Blum, Indiana University.

Consider an arbitrary sequence of random variables $X_1, X_2, \cdots$, to be observed sequentially and a corresponding sequence of statistics $f_1(X_1)$, $f_2(X_1, X_2), \cdots$, $f_j(X_1, \cdots X_j), \cdots$ each of the latter with zero expectations. With m denoting the number of random variables observed, determined by a sequential stopping rule, a necessary and sufficient condition that $E(f_m) = 0$ for all truncated sequential stopping rules is that the sequence $f_1, f_2, \cdots$ be a martingale. Applications of this result is made to expectations of sums and products including a form of the fundamental identity of sequential analysis which is valid for unbounded stopping rules.

2. Exact Nonparametric Tests for Randomized Blocks. JOHN E. WALSH, Systems Development Corporation, Santa Monica, California. (By title)

A class of nonparametric procedures for testing the statistical identity of treatments in randomized block experiments is suggested and discussed. The suggested procedures are squarely based on experimental within-block randomizations, and they may be chosen so as to have special power against particular alternatives. The blocks are assumed to be statistically independent but no assumption is made concerning the dependence within the various blocks. The basic idea is to obtain from each block a statistic that is, under the null hypothesis, symmetrically distributed about zero and then apply a nonparametric test of symmetry about zero. The observational data can be of any quantitative type.

3. On the Determination of Joint Distributions from the Marginal Distributions of Linear Combinations. THOMAS S. FERGUSON, University of California, Los Angeles.

Let $Z_n = \alpha_n X + \beta_n Y$ where $\gamma_n = \alpha_n/\beta_n$ are all distinct for $n = 1, 2, \cdots$. A sufficient condition that the joint distribution of (X, Y) be determined uniquely by the distributions of the Z_n is that there exist an integer m such that (1) $E \exp\{t|Z_m|\} < \infty$ for some $t > 0$ and (2) there is a limit point of the γ_n's (possibly $\pm\infty$) other than γ_m. Conversely when the joint distribution has a piece of positive continuous density somewhere, the distributions of a finite number of the Z_n's do not determine the joint distribution. Thus in particular the bivariate normal distribution is determined uniquely by any infinite collection of distinct linear combinations of the variables and by no finite number of them. These results extend to many dimensions, with suitable modifications.

4. Approximation to the Probability Density of Zero—Crossings Intervals of a Gaussian Process. SYLVAIN EHRENFELD, New York University. (By title)

Let $x(t)$ be a stationary Gaussian process with a given spectrum $w(f)$, and let $P_0(t)$ be the probability density of the lengths of intervals between successive zeros in this process. In the present paper several approximations to $P_0(t)$ are obtained. This is achieved by the evaluation of multiple integrals whose evaluation are equivalent to finding

"volumes" of hyperspherical simplices in n-dimensional space. The above geometrical problem is solved for certain cases ($n = 3, 4$) and given the desired approximations. The problem of finding the volume of a hyperspherical simplex reduces (for $n = 4$) to the problem of finding the 3-dimensional volume of a tetrahedron-like figure (the analogue of a spherical triangle) on the surface of a 4-dimensional sphere. The final part of the paper is concerned with a comparison between the various approximations to a count of zero crossings, taken from a series of ocean pressure records. There also is a discussion of computational problems.

5. The Probability in the Extreme Tail of a Convolution. DAVID BLACKWELL AND J. L. HODGES, JR., University of California, Berkeley. (By title)

Let $X_1, X_2, \cdots$ be independent identically distributed integer-valued random variables, such that 0 is a possible value and the g.c.d. of possible values is 1. Suppose $E(e^{tX_1})$ is finite for some $t > 0$. For any number a with $E(X_1) < a < \sup X_1$ there are a distribution ρ on the values of X_1 and a number m, $0 < m < 1$ such that Prob $\{X_1 + \cdots + X_n = na\} = \pi_n^{**}(1 + 0(n^{-1}))$ as $n \to \infty$ through those values for which na is a possible value of $X_1 + \cdots + X_n$, where $\pi_n^{**} = m^n[1 + ((\mu_4/\mu_2^2) - 3 - (5\mu_3^2/3\mu_2^3))1/8n]/\sigma\sqrt{2\pi n}$, and $\sigma, \mu_2, \mu_3, \mu_4$ are the standard deviation and central moments of the ρ distribution. By ignoring certain error terms, an approximation Prob $\{X_1 + \cdots + X_n \geqq na\} = c\pi_n^{**}(1 - d/n)(1 + 0(n^{-1}))$ is obtained. It is noted that Prob $\{X_1 + \cdots + X_n = na\} = m^n$ Prob $\{Y_1 + \cdots + Y_n = na\}$, where $Y_1, Y_2, \cdots$ are independent variables with the ρ distribution. Some numerical illustrations of the accuracy of the approximations are given.

6. Asymptotic Methods of Evaluating the Integral from a to ∞ of $f(x)$. WYMAN RICHARDSON, University of North Carolina. (By title)

Three iterative procedures are considered. Procedure A: a transformation is applied carrying ∞ into 0 and a into b. Then the integral, "$F(a)$", is expanded in a Taylor series about 0 or b. Procedure B: $F_2(a) = F_1(a)f_1'(a)/f_1(a)$, where $f_1(a) = -F_1'(a)$. Procedure C: $F(a) = f(a)[v_1(a) + \cdots + v_n(a)] + \int_a^\infty (x)v_n'(x)dx$, where $v_1(a) = -f(a)/f'(a)$ and $v_n(a) = v_1(a)v_{n-1}'(a)$, (Laplace, Winckler). A battery of order theorems are proved, using the Cauchy definition: "order of $f = r$" means "$x^{-r-\epsilon} < f(x) < x^{-r+\epsilon}$ for x large". The order of the "relative error", $F_1(x)/F(x) - 1$, equals that of the "relative frequency error", $f_1(x)/f(x) - 1$. If the order of $f = \infty$, B and C are usually "asymptotic I": the order of the relative error $\to \infty$ as the number of terms increases; and "asymptotic II": (the relative error for $n + 1$ terms)/(the relative error for n terms) $\to 0$. If the order of f is finite, linear combinations of the terms are taken to make the procedure asymptotic II. Useful formulae, such as $F(a) \cong -r[f(a)]^2/f'(a)(r - 1)$, and $-[f(a)]^2/f'(a)[2 - u(a)]$, where $u(x) = f(x)f''(x)/[f'(x)]^2$, are obtained from a few terms of these procedures and applied to statistical distributions. They are asymptotic as $a \to \infty$. Analogous procedures for sums of series and finite tails are considered.

7. Some Properties of Binary Arrays Which Are Generated by Iterated Sequences and Reversals. H. VON GUERARD, Lockheed Aircraft Corporation.

From a sufficiently extended binary event another one is deduced, by assigning 0's to sequences and 1's to reversals. This operation satisfies the group of additivity of a binary (modulation 2) ring algebra (R. D. Bose and R. R. Kuebler, Univ. N.C., Inst. Stat., Mimeo. Ser. 199 (1958)). Repeated application generates an array of binary numbers, whose fre-

quency ratios (= No. of 1's/No. of 0's) (a) converge to 1, (b) repeat periodically, (c) assume relative extrema by geometric progression of iteration order, (d) behave irregularly, i.e. neither (a), (b) nor (c),—if, and only if, the initiating process is (a) randomized, incl. conditioned processes, (b) periodic, (c) transient (single peaks or steps), (d) inductive. To (b): M. Kochen and E. H. Galanter (*Inform. Contr.* 1, 267–288 (1958)) determined the elements of minimal generating sets (mgs) for λ-placed binary numbers (here: periods), from which all others could be deduced by completion or by translation. By iterated addition (mod. 2), any periodic process generates periodically repeated arrays (of periodic processes), whose elements are exactly those of the mgs's, none of them occurring more than once. Since these arrays, in the average, imply more than 1 element, as λ increases, their number becomes progressively small if compared to the number of elements of the mgs's.

(Additional abstract for the Cambridge Meeting of the Institute, August 25–29, 1958)

30. **On the Bounds for the Variance of Mann-Whitney Statistic.** J. S. RUSTAGI, Michigan State University (By title)

Let $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ be two random samples from strictly increasing continuous cumulative distribution functions (cdf's) $F(x)$ and $G(y)$ respectively. Then the Mann-Whitney statistic U is given by U = number of pairs (X_i, Y_j) such that $Y_j < X_i$, $i = 1, 2, \cdots, m$; $j = 1, 2, \cdots, n$. Let $L(t) = F(G^{-1}(t))$. Then variance of U, $V(U) = mn[(m-1)\int_0^1 (L(t) - kt)^2\,dt + A]$ where A is a constant free of $L(t)$ and $k = (n-1)/(m-1)$. Utilizing the techniques and results of an earlier paper by the author (*Ann. Math. Stat.*, Vol. 28, pp. 309–328), lower and upper bounds for $V(U)$ are determined in terms of $p = P(Y < X) = \int_{-\infty}^{\infty} F(t)\,dG(t) = 1 - \int_0^1 L(t)\,dt$. The problem essentially is that of minimizing and maximizing $\int_0^1 (L(t) - kt)^2\,dt$ over a class of cdf's $L(t)$ defined over [0, 1] such that $\int_0^1 L(t)\,dt = 1 - p$. Lower bounds are also obtained for $V(U)$ under an additional restriction that X is stochastically smaller than Y or $L(t) \geq t$ for $0 \leq t \leq 1$. (Received July 7, 1958, revised November 24, 1958).

NEWS AND NOTICES

Readers are invited to submit to the Secretary of The Institute news items of interest

Personal Items

Frances Campbell Ameniya, formerly chairman of the Department of Mathematics at George Pepperdine College in Los Angeles, California, has been appointed Associate Professor of Mathematics at California Western University in San Diego, California.

R. E. Barlow is now working on a doctorate at Stanford while employed at Sylvania Electronic Defense Laboratory, Mt. View, California, as a mathematical statistician.

Ishu Bangdiwala is on a leave of absence from his position as Head of the Department of the Statistics Section of the Agricultural Experiment Station of the University of Puerto Rico, to accept the position as Assistant Director of Research in the Superior Council on Education, which is the governing board of the University.

Jerome Cornfield, assistant chief of the Biometrics Branch, Division of Re-

search Services, National Institute of Health, has been appointed to two professorships at the Johns Hopkins Medical Institutions. He is professor and chairman of the Department of Biostatistics in the School of Hygiene and Public Health, succeeding William G. Cochran, and also fills the newly created post of professor of Biomathematics in the School of Medicine.

John W. Cotton has returned to his post as assistant professor of Psychology at Northwestern University following a year as Postdoctoral Fellow in the Department of Statistics at the University of Chicago.

Jean Engler is spending her second year of a postdoctoral National Science Foundation Fellowship at the Department of Statistics, Harvard University.

Edgar H. Fickensher, formerly Principal of Oroville Union High School, Oroville, California, has moved to Stanford, California to finish work on a Ph.D. in Education.

Paul Gunther is currently working as a consultant in statistics and operations research.

Dr. Shanti S. Gupta spent the academic year 1957–58 as an associate professor in the Mathematics Department of the University of Alberta and later was a fellow at the Summer Research Institute of the Canadian Mathematical Congress at Queen's University. He has now joined the Bell Telephone Laboratories in Allentown, Pennsylvania as a member of the Technical Staff.

Dr. John W. Hamblen, formerly Associate Professor of Mathematics and Director of the Computing Center at Oklahoma State University, has been employed as full time Director of the Computing Center which has been established under the general supervision of the Office of the Vice President of the University of Kentucky.

Nancy Lee Hannye is now Assistant Professor at Michigan State University.

Dr. Bernard Harris has resigned his position as Mathematician, U. S. Department of Defense and accepted an appointment as Assistant Professor of Mathematics, University of Nebraska.

Assistant Professor L. L. Helms of Michigan State University has been appointed to an assistant professorship at the University of Illinois.

David Hogben has obtained a leave of absence from the Western Electric Co., Kearny, New Jersey, to pursue studies at Iowa State College for the Ph.D. degree.

John R. Howell has recently joined the staff of the University of Dayton Research Institute as head of the Computer Section.

Howard L. Jones has retired as General Supervisor of Statistics with the Illinois Bell Telephone Company, and has accepted an appointment as Professor of Statistics in the School of Business at the University of Chicago.

Dr. Orval M. Klose, formerly Associate Professor and Head of the Mathematics Department at Seattle University, has accepted a position as Associate Professor of Mathematics at Humboldt State College in Arcata, California.

John M. Leiman has resigned his position with the Personnel Laboratory,

Wright Air Division Center, and has accepted a position with System Development Corporation of Santa Monica, California.

Dr. Radha G. Laha has been appointed Research Assistant Professor in the Department of Mathematics at the Catholic University of America for the year 1958–59.

Ronald Pyke has accepted a position as Assistant Professor in the Department of Mathematical Statistics at Columbia University for the year 1958–59. He formerly held a similar position at Stanford University.

Philburn Ratoosh, on leave from the Ohio State University, is Visiting Associate Professor of Psychology, at the University of California, Berkeley, for 1958–59.

Enders A. Robinson has been appointed Assistant Professor in the Department of Mathematics at the University of Wisconsin, Madison, Wisconsin.

Jagdish S. Rustagi, formerly with the Department of Statistics, Michigan State University, has accepted the position of Reader in Statistics, Department of Mathematics, Muslim University, Aligarh, India.

Professor Henry Scheffé of the University of California will be Visiting Professor at Princeton University for the academic year 1958–59, where his work with the Statistical Techniques Research Group will be partially supported by a grant from the National Science Foundation.

After spending the 1957–1958 year as a Fellow at the Center for Advanced Study in the Behavioral Sciences, Sidney Siegel has returned to his position as Associate Professor of Psychology at Pennsylvania State University.

Jack Silber spent the summer of 1958 as Consultant to the Assistant for Operations Analysis, USAF and has returned to Roosevelt University as Professor of Mathematics and Acting Assistant to the Dean of Faculties.

Romuald Slimak has been recently appointed Manager, Univac Computing Center, Remington Rand Univac Division, Sperry Rand Corporation, 315 4th Avenue, New York 10, New York.

Seiji Sugihara is now Research Specialist at Lockheed Aircraft Corporation, Missile Systems Division, Sunnyvale, California.

Dr. Balkrishna V. Sukhatme has been appointed Professor of Statistics in the Indian Council of Agricultural Research, New Delhi, India.

Dr. John W. Wilkinson, Formerly Assistant Professor of Mathematics at Qeuen's University, Kingston, Ontario, has accepted a position as Research Statistician at Westinghouse Research Laboratories, Pittsburgh 35, Pennsylvania.

Dr. W. H. Williams has resigned as Assistant Professor of Statistics at Iowa State College to become Assistant Professor of Mathematics at McMaster University, Hamilton, Canada.

Mr. D. M. G. Wishart has been appointed Lecturer in Statistics in the Department of Pure Mathematics of the University of Birmingham, England, from October 1, 1958.

New Members

The following persons have been elected to membership in the Institute

June 23, 1958, to October 16, 1958

Adorno, David S., M.A. (Penn. State Univ.), Research Assistant, Dept. of Statistics, Harvard University, Cambridge, Mass.; *323 South Road, Bedford, Mass.*

Ando, Albert K., Ph.D. (Carnegie Inst. of Tech.), Assistant Professor, *Dept. of Economics, Massachusetts Institute of Technology, Cambridge 39, Mass.*

Bagai, Om Parkash, M.A. (Panjab, India), Student for Ph.D., Math., U.B.C., Vancouver, Canada; *Dept. of Math., U.B.C., Vancouver, Canada.*

Bardwell, George E., M.S. (Univ. of Colorado), Assistant Professor, University of Denver; *1445 Cleveland Place, Denver 2, Colorado.*

Barney, Jesus, C.P.A. (Monterrey Inst. of Tech.), Professor of Math. Stat. and Auditing, Monterrey Institute of Technology, Monterrey, N.L., Mexico; *Sucursal De Correos "J", Monterrey, N.L., Mexico.*

Cannon, L. Dennis, M.S. (Purdue Univ.), Graduate Research Assistant, Purdue University Statistical Laboratory, West Lafayette, Indiana; *137 South Salisbury St., West Lafayette, Indiana.*

Chacko, V. John, M.Sc. (Univ. of Trawaneore), Graduate Student, University of California (T.A.), University of California, Berkeley; *Dept. of Statistics, University of California, Berkeley 6, California.*

Chipman, John S., Ph.D. (The Johns Hopkins Univ.), Associate Professor of Economics, University of Minnesota, Minneapolis 14, Minn.; *Dept. of Economics, University of Minnesota, Minneaplis 14, Minn.*

Christie, Theodore J., B.S., (Rutgers Univ.), Doctoral Candidate, Dept. of Ind'l. and Admin. Eng'g., Cornell University, Ithaca, New York; *78 Roosevelt Street, Cresskill, New Jersey.*

Cogburn, Robert F., A.B. (Univ. of California), Student, University of California, Berkeley 4, California; *1716A Francisco, Berkeley 3, California.*

Copeland, Lewis C., Ph.D. (Duke Univ.), Acting Head of Dept. and Professor of Statistics, *Dept. of Statistics, University of Tennessee, Knoxville, Tenn.*

Corsten, Leo C. A., Dr. (Agr. Univ., Wageningen, Netherlands), Head, Statistical Dept., IVRO, Wageningen, Netherlands; *400 Smith Ave., Chapel Hill, North Carolina.*

Current, James L., MAT (Indiana Univ.), Mathematician, National Security Agency, Washington 25, D. C.; *4521 29th Street, Mt. Rainier, Maryland.*

De Figueiredo, Djairo Guedes, C.E. (National School of Engineering, Rio de Janeiro, Brazil), fellowship from National Council of Researches (Brazil), New York University; *102-22 62nd Road, Forest Hills, 75, N. Y.*

Folks, John Leroy, Ph.D. (Iowa State College), Operations Research Engineer, Texas Instruments Incorporated, 13500 N. Central Expressway, Dallas, Texas; *P. O. Box 312, Dallas, Texas.*

Goldman, Herbert M., M.A. (Univ. of Illinois), Statistician, University of Illinois, Urbana, Illinois; *1111 S. Arbor Ave., Champaign, Ill.*

Foradori, George Thomas, M.S. (N. C. State College), Research Assistant, *Dept. of Experimental Statistics, N. C. State College, Raleigh, North Carolina.*

Gabriel, Kuno Ruben, Ph.D., (Hebrew Univ.), Instructor, *Hebrew University, Jerusalem, Israel.*

Hager, Frederick W., MS (Univ. of Delaware), Assistant Professor of Mathematics, *United States Naval Academy, Annapolis, Maryland.*

Hawthorne, George Boltz, Jr., M.S (Georgia Inst. of Tech.), Assistant Professor of E. E. and part-time student, *Georgia Institute of Technology; 225 North Avenue, Atlanta 13. Georgia.*

Hefner, Oscar V., B.S. (Georgia Inst. of Tech.), Graduate Student, Georgia Tech.; *Rich Electronic Computer Center, Ga. Tech., Atlanta 13, Ga.*

Hernandez, Antonio, Ingeniero Industrial, (Escuela Tecnica Superior de I.I.), Professor of Fundamental Statistics and Assistant to the Managing Director of G. E. Espanola, Escuela Tecnica Superior de I.I. and General Electrica Espanola, S.A., Campo de San Mames, Plaza de Federico Moyua 4, Bilbao, Spain; *Egana 14, Bilbao, Spain.*

Johnson, Linwood A., B.I.E., (Georgia Inst. of Tech.), Graduate Student, Georgia Institute of Technology, North Avenue, Atlanta, Ga.; *281 N. Colonial Homes Circle, N. W., Atlanta 9, Ga.*

Klotz, Jerome H., A.B., (Univ. of California), Research Assistant, University of California, Berkeley, California; *Dept. of Statistics, Univ. of California, Berkeley 4, California.*

Koenig, Robert A., M.S. (Rutgers Univ.), Statistician, National Lead Co., Titanium Division, P. O. Box 58, South Amboy, New Jersey; *4 Maple Avenue, Matawan, New Jersey.*

Larson, Harold J., M.S. (Iowa State College), General Electric Fellow in Statistics, Iowa State College; *Statistical Laboratory, Iowa State College, Ames, Iowa.*

Littell, Arthur S., Sc.D. (Johns Hopkins Univ.), Assistant Professor of Biostatistics, *Western Reserve University, School of Medicine, Cleveland 6, Ohio.*

McGahey, Mary B., B.S. (Radford College), Graduate Student in Statistics, V.P.I. Blacksburg, Va.; *Box 11—Station A, Radford, Virginia.*

Mallios, William S., B.S. (Purdue Univ.), Research Assistant, N. C. State College; *Box 5457, Raleigh, North Carolina.*

Maloney, Richard C., M.S., (Univ. of Southern California), Supervisor, Statistical Analysis Unit, Reliability Gp., Rocketdyne, 6633 Canoga Blvd., Canoga Park, California; *7112 Delco Ave., Canoga Park, California.*

Maxwell, William L., B.M.E. (Cornell Univ.), Graduate Student, Cornell University, Ithaca, New York; *Upson Hall, Cornell University, Ithaca, New York.*

Mendelsohn, Jay, B.A., (New York Univ.), Research Engineer, Grumman Aircraft Engineering Corporation, Bethpage, Long Island, New York; *150-40 71st Avenue, Flushing 67, New York.*

Mikhail, Wadie Fultas, M.Sc. (Univ. of Cairo, Egypt), Graduate Student, Dept. of Statistics, Univ. of North Carolina; *210 Phillips Hall, Chapel Hill, North Carolina.*

Neathammer, Robert D., M.A. (Univ. of Illinois), Mathematician, U. S. Navy Ammunition Depot, Quality Evaluation Laboratory, Crane, Indiana; *518 Cedar Street, Centralia, Illinois.*

Olson, Milton P., B.S. (Stanford Univ.), Student, Stanford University; *405 Park Street, Turlock, California.*

Patil, Ganapati P., M.Sc. (Univ. of Poona, India), Teaching Fellow, *Dept. of Mathematics, University of Michigan, Ann Arbor, Michigan.*

Ray, Santosh Kumar, M.Sc. (Lucknow Univ., India), Graduate Student and Research Assistant, *Dept. of Mathematical Statistics, Columbia University, New York 27, New York.*

Rhodes, Benjamin T., Jr., M.A. (Univ. of Texas), Graduate Assistant, Oklahoma State University, Stillwater, Oklahoma; *Unit 31, Apt. 1; North University Place, Stillwater, Oklahoma.*

Rivers, Joye Boring, MEd., (Univ. of Florida), Research Assistant and Student, Radar Development Branch, Engineering Experiment Station, Georgia Institute of Technology, 225 North Ave., Atlanta 13, Ga.; *4785 E. Conway Dr., N.W., Atlanta 5, Georgia.*

Rothstein, Marvin, M.S. (New York Univ.), Mathematician, Service Bureau Corp., 635 Madison Ave., New York, N. Y.; *66 West 53rd Street, New York, N. Y.*

Roy, Jagabrata, D.Phil. (Univ. of Calcutta), Lecturer, *Research and Training School, Indian Statistical Institute, 203 B.T. Rd., Calcutta 35, India.*

Shuford, Emir H., Jr., Ph.D. (Univ. of Illinois), Assistant Professor, Dept of Psychology, University of North Carolina, Chapel Hill, N. C.; *Psychometric Laboratory, University of North Carolina, Chapel Hill, N. C.*

Siniff, Donald B., M.S. (Michigan State Univ.), 2nd Lt., USAF, Student Mail Room, Box 5840, Harlingen AFB, Harlingen, Texas; *Box 432, Grafton, Ohio.*

Sowinski, John J., S.B. (Depaul Univ.), Research Statistician, The Toni Company, Division of the Gillette Company, *456 Merchandise Mart, Chicago, Ill.*

Starr, Selig, M.A., (George Washington Univ.), Head, Applied Mathematics and Statistics Group, Research and Development Dept., U.S. Naval Propellant Plant, Indian Head, Maryland; *812 University Blvd., E., Silver Spring, Md.*

Stiassny, Simon, M.A., (Univ. of California), Statistician, I.B.M. Research Center, *Box 218, Yorktown Heights, New York.*

Sullivan, Rebecca J., B.A. (Michigan State Univ.), Graduate Research Assistant, Dept. of Statistics, Michigan State University, East Lansing, Michigan; *2209 Alpha St., Lansing, Mich.*

Tsao, Rhett F.S., M.A. (Oklahoma State Univ.), Graduate Assistant in Statistics, Oklahoma State University, Stillwater, Oklahoma; *Statistical Laboratory Dept. of Mathematics, Oklahoma State University, Stillwater, Oklahoma.*

Vithayasai, Chitra, B.S. (Chulalongkorn Univ. of Thailand), Rice Statistician, Rice Dept., Ministry of Agriculture, Bangkok, Thailand; *Dept. of Plant Breeding, Cornell University, Ithaca, New York.*

Wagner, Harvey, M., M.S. (Stanford Univ.), Assistant Professor, *Industrial Engineering and Statistics, Stanford University, Stanford, California.*

West, Del L., B.S. (Southeastern State College), Graduate Assistant, Oklahoma State University, Stillwater, Oklahoma; *645 Bennett Dr., Stillwater, Okla.*

Wheatley, Thomas V., M.S. (Illinois Inst. of Tech.), Quality Control Engineer, Convair, A Division of General Dynamics, Pomona, California; *P. O. Box 884, Pomona, California.*

Woo, Jae Lin, M.S. (Massachusetts Inst. of Tech.), Instructor, Textile Eng'g., College of Engineering, Seoul National University, Seoul, Korea; *403 Beacon Street, Box 15, Massachusetts.*

Woo, Juo Chuan, M.S. (N. C. State College), Assistant Director of Processing Research, *The Textile Research Center, School of Textiles, N. C. State College, Raleigh, North Carolina.*

Yamane, Taro, Ph.D., (Univ. of Wisconsin), Assistant Professor, *New York University, School of Commerce, Dept. of Economics, Washington Square, New York 3, New York.*

Yamamoto, Sumiyasu, M.S. (Hiroshima Univ.), Professor of Statistics, *Dept. of Statistics, Nara Medical College Kashihara, Nara, Japan.*

Zadeh, Lofti A., Ph.D. (Columbia Univ.), Professor of Electrical Engineering, Columbia University, New York 27, New York; *850 James Street, Pelham Manor, New York.*

Zilmer, Delbert E., Ph.D. (Univ. of Wisconsin), Mathematician, U.S. Naval Ordnance Test Station, China Lake, California; *701-B Lexington Ave., China Lake, California*

SUMMER OFFERINGS IN STATISTICS AT IOWA STATE COLLEGE

The Department of Statistics at Iowa State College will offer six applied courses in statistical theory and methods in its two 1959 summer sessions. These courses are planned primarily for graduate students or research workers with limited mathematical backgrounds who wish to use statistical techniques intelligently for application to other fields. In addition, a course on special topics in theoretical or applied statistics may be studied at the graduate level. Senior staff members will be available during most of the summer for consultations on research or special problems.

Students may register for either or both of the six-week summer sessions: June 8–July 15 and July 15–August 21. The complete list of statistics offerings for the first session is as follows: Stat. 401, "Statistical Methods for Research Workers" (at the level of Snedecor's *Statistical Methods*): Stat. 447, "Statistical Theory for Research Workers" (mainly theory of experimental statistics at the level of Anderson and Bancroft's *Statistical Theory in Research*); Stat. 599, "Special Topics"; and Stat. 699, "Research." In the second session will be offered Stat. 402, a continuation of 401; Stat. 448, a continuation of 447; two courses in applied methods which are more specialized, Stat. 411, "Experimental Designs for Research Workers," and Stat. 421, "Survey Designs for Research Workers"; and finally Stat. 599 and 699. Additional information may be obtained from T. A. Bancroft, Department Head and Director, Statistical Laboratory, Iowa State College.

INTERDISCIPLINARY CONFERENCE ON SELF-ORGANIZING SYSTEMS

An Interdisciplinary Conference on Self-Organizing Systems will be held on May 5th and 6th, 1959, at the Museum of Science and Industry, Chicago, Illinois. The conference is to be co-sponsored by the Information Systems Branch of the Office of Naval Research and the Armour Research Foundation. The purpose of this conference is to bring together research workers in all fields of science who are concerned either with the development of self-adaptive information systems or with the conduct of research which may contribute to an improved understanding of cognitive, learning, and growth processes. Particular emphasis will be placed on theoretical models of systems which are capable of spontaneous classification, identification, and symbolization of their inputs.

Interested individuals may receive further information and a preliminary conference program when available by writing to Mr. Scott Cameron, ICSOS Conference Secretary, Armour Research Foundation, 10 West 35th Street, Chicago 16, Illinois.

A NEW JOURNAL OF STATISTICS

The Statistical Society of New South Wales, founded in 1947, is the only Society in Australia concerned with all aspects of statistics and related fields. Among its activities are general and section meetings and symposia. The latter have been attended by people from all over Australia. It has also published a Bulletin; the present editor is H. S. Konijn.

The Society has now been enabled to issue a printed Journal, to be called *The Australian Journal of Statistics*, and to be issued three times a year. The Editorial Board consists of H. S. Konijn, H. O. Lancaster and R. S. G. Rutherford (address: The University of Sydney, Sydney, Australia). The Journal expects to have contributions from Australia and abroad in the field of statistical theory and applications. Prospective authors are requested to send for a list of

instructions regarding the form of the manuscripts. Issues will be available to non-members at 10 shillings a copy.

NATIONAL SCIENCE FOUNDATION BULLETINS

The National Science Foundation announces a new series of bulletins which, when completed, will represent an inventory listing of all significant scientific information sources or activities within the Federal Government. The primary objective of this program is to make unclassified, unpublished scientific research information easily accessible and readily available to all U. S. scientists and engineers, both in and out of the Government.

An earlier NSF report, "Organization of the Federal Government for Scientific Activities", describes the pattern according to which the various Government agencies are organized for the conduct of basic research, applied research, development, and other scientific activities including scientific information. The documents in this series will present, for these same Government branches, information on general subject fields in which scientific reports are prepared, categories of scientific reports issued, policies regarding the announcement and availability of these reports to the scientific community, and locations and policies of the agencies' libraries and information centers, etc.

Bulletin No. 1, October 1958, gives information on the Department of Agriculture. Further information may be obtained from the URI Program Director, National Science Foundation, Science Information Service, Washington 25, D. C.

VISITING FOREIGN MATHEMATICIANS

The following list (dated October 27, 1958) of visiting foreign mathematicians has been received from the Division of Mathematics, National Academy of Sciences, National Research Council. The information given is, in order, the name, home country, host institution, and period of visit; AY stands for academic year. The names of persons whose visit terminates before May, 1959, have not been included.

Alvarez de Araya, Jorge, Chile, University of Washington, Sept. 1958–Sept. 1959; Austin, M., U. K., University of Wisconsin, Sept. 1958–Sept. 1959; Baayen, Pieter C., Netherlands, University of California, Berkeley, AY 1957–59; Besicovitch, A. S., England, University of Pennsylvania, Sept. 1958–June 1959; Bialynicki, Andrzej, Poland, University of California, Berkeley, AY 1958–59; Blakers, Albert, Australia, Princeton University, Sept. 1958–August 1959; Bofinger, Victor, Australia, North Carolina State College, June 1957–June 1959; Chapman, Sydney, England, University of Michigan, 1958–59; Chartres, Bruce A., Australia, Brown University and Massachusetts Institute of Technology, Sept. 1958–May 1959; Chwe, Byoung-Song, Korea, University of California, Berkeley, AY 1957–59; Craggs, James W., U. K., Brown University, Sept. 1958–Sept. 1959; Drazin, Philip G., England, Massachusetts Institute of Technology, Sept. 16, 1958–June 15, 1959; Duguid, A. M., Australia, Brown University, Aug. 15 1958–May 31, 1959; Flon,

PETER, Australia, Duke University, July 1958–June 1959; FOGUEL, SHAUL R., Israel, University of California, Berkeley, Sept. 1958–June 1959; FUHRKEN, GEBHARD, Germany, University of California, Berkeley, AY 1957–59; GANI, JOSEPH M., Australia, Columbia University, Feb. 1959–Feb. 1960; GREWE, RUDOLF, Spain, University of California, Berkeley, AY 1957–59; GRIBBEN, RONALD J., England, Massachusetts Institute of Technology, Sept. 16, 1958–June 15, 1959; HA, KWANG CHUL, Korea, University of North Carolina, Sept. 15, 1958–June 30, 1959; HANO, JUN-ICHI, Japan, University of Chicago, Sept. 1, 1958–Aug. 31, 1959; HELMBERG, GILBERT M., Austria, Institute of Int. Education and Tulane University, July 1958–Sept. 1959; HELMS, HANS, Denmark, Tufts University, Sept. 1958–June 1959; HILTON, PETER, England, Cornell University, Sept. 1958–Sept. 1959; HULL, T. E., Canada, California Institute of Technology, Sept. 1, 1958–Aug. 31, 1959; IRMAY, SHRAGGA, Israel, Institute of Mathematical Sciences, New York University, Sept. 1958–August 1959; KLINGEN, HELMUT, Germany, University of California, Sept. 1958–June 1959; KRISTENSEN, LEIF, Denmark, Yale University, Aug. 1958–Aug. 1959; KRZYWICKI, A., Poland, University of Kansas, Oct. 1958–July 1959; LAHA, RADHA G., India, Catholic University, Sept. 1957–May 1959; LEIS, ROLF, Germany, Institute of Mathematical Sciences, New York University, Sept. 1958–Aug. 1959; LEVY, AZRIEL, Israel, Massachusetts Institute of Technology, Sept. 16, 1958–June 15, 1959; MATUSITA, KAMEO, Japan, Princeton University, AY 1958–59; MITCHELL, A. R., Scotland, California Institute of Technology, Mar. 1, 1959–Sept. 30, 1959; MORLAND, L. W., England, Brown University, Sept. 1, 1958–May 31, 1959; MOSTOWSKI, ANDRZEJ, Poland, University of California, Berkeley, Sept. 1958–June 1959; MUKI, R., Japan, Brown University, Sept. 15, 1958–June 30, 1959; MUNN, W. D., Scotland, Tulane University, Sept. 1958–Sept. 1960; NAGAO, HIROSI, Japan, University of Michigan, 1958–1959; NAGATA, MASAYOSHI, Japan, Harvard University, Sept. 1, 1958–June 30, 1959; NANNINI, AMOS, Italy, University of Minnesota, Sept. 1958–June 1959; OBATA, MORIO, Japan, University of Illinois, Sept. 1958–Sept. 1959; ODEH, FAROUK J. S., Jordan, University of California, Berkeley, AY 1956–59; OGAWA, JUNJIRO, Japan, University of North Carolina, August 1958–June 30, 1959; OHTSUKA, M., Japan, University of Kansas, March 1959–May 1959; OIKAWA, KOTARO, Japan, University of California, Los Angeles, July 1, 1958–June 30, 1959; O'KEEFFE, JEREMIAH, Ireland, Institute of Mathematical Sciences, New York University, Sept. 1958–August 1959; OKUBO, TANJIRO, Japan, State College of Washington, Sept. 16, 1958–June 10, 1959; PEDERSEN, FLEMMING PER, Denmark, University of Southern California, July 1958–July 1959; PLEIJEL, A., Sweden, University of Kansas, March 1959–May 1959; PLIS, ANDRZEJ, Poland, Institute of Mathematical Sciences, New York University, Sept. 1958–Aug. 1959; PROUDMAN, IAN, U. K., California Institute of Technology, Sept. 1958–Sept. 1959; ROHRL, HELMUT, Germany, University of Chicago, Oct. 1, 1958–June 30, 1959; RUBEN, HAROLD, England, Columbia University, AY 1957–59; RUTOVITZ, DENIS, So. Africa, University of California, Berkeley, Sept. 1958–June 1959; SABHARWAL, RANJIT SINGH, India, University of California, Berkeley, AY 1958–59; SABIDUSSI, GERT O., Austria, Tulane University, Sept. 1955–Sept. 1959; SAKURAI, AKIRA, Japan, Institute of Mathematical Sciences, New York University, Sept. 1958–Aug. 1959; SCHAEFER, HELMUT, W. Germany, State College of Washington, Mar. 19, 1958–July 1959; SCHMIDT, PALLE FROSIG, Denmark, University of Minnesota, Sept. 1958–Sept. 1959; SCHOFF, ANDREAS, Switzerland, American University and National Bureau of Standards, Oct. 1957–June 1959; SCRIBA, J. CRISTOPH, Germany, University of Massachusetts, Aug. 1958–Sept. 1959; SHISHA, OVED, Israel, Technion-Israel Institute, Sept. 1, 1958–June 30, 1959; SIOSON, FREDERICO M., Philippines, University of California, Berkeley, AY 1956–59; SKOVGAARD, H., Denmark, California Institute of Technology, Jan. 1, 1959–June 30, 1960; SZARSKI, J., Poland, University of Kansas, Oct. 1958–July 1959; VERMA, G. R., India, Institute of Mathematical Sciences, New York University, Sept. 1958–August 1959; VOGEL, WALTER, Germany, University of Chicago, Oct. 1958–June 1959; WALLACE, DAVID, U. K., Princeton University, Sept. 1958–Sept. 1959; WALTER, WOLFGANG, Germany, University of Maryland, Sept. 1, 1958–June 30, 1959; WATSON, GEOFFREY, Australia, Princeton University,

Sept. 1958–June 1959; WONG, YUNG-CHOW, Hong Kong, Institute for Advanced Study, Sept.–Dec. 1958, University of Chicago, Feb. 1–June 30, 1959; Sept. 1958–June 30, 1959; WOODS, A. C., England, Tulane University, Sept. 1957–Sept. 1959; YASUHARA, MITSURU, Japan, University of California, Berkeley, AY 1957–59; YEVDJEVICH, V. M., Yugoslavia, American University and National Bureau of Standards, Feb. 1958–June 1959; YOUNG, EUTIQUIO, Philippines, University of Maryland, Sept. 1958–Sept. 1959; ZADUNAISKY, PEDRO, Argentina, Princeton University, 1957–1959; ZASSENHAUS, H. J., Canada, California Institute of Technology, Sept. 1, 1958–Aug. 31, 1959.

SOUTHERN REGIONAL GRADUATE SUMMER SESSION IN STATISTICS AT NORTH CAROLINA COLLEGE, 1959

The 1959 session of the Southern Regional Graduate Summer Session in Statistics will be held at North Carolina State College, Raleigh, from June 8 to July 17, 1959.

North Carolina State College, Virginia Polytechnic Institute, University of Florida, and Oklahoma State University have agreed to operate a continuing program of graduate summer sessions in statistics to be held at each institution in rotation; the program was instituted at Virginia Polytechnic Institute in the summer of 1954.

The 1959 session, like previous sessions under this program, is intended to serve: 1) teachers of introductory statistical courses who want formal training in modern statistics; 2) research and professional workers who want intensive instruction in basic statistical concepts and modern statistical methodology; 3) professional statisticians who wish to keep informed about advanced specialized theory and methods; 4) prospective candidates for graduate degrees in statistics; and 5) graduate students in other fields who desire supporting work in statistics.

The session will last six weeks and courses will carry three semester hours of credit. Not more than two courses may be taken for credit at any one session. The summer work in statistics may be applied as residence credit at any of the cooperating institutions, as well as certain other universities, in partial fulfillment of the requirements for a graduate degree. The program may be entered at any session, and consecutive courses will follow in successive summers so that it would be possible for a student to complete the course work in statistics for a Master's degree in three summers. Students must satisfy the remaining requirements for course work and thesis at the institution where they are to be admitted to candidacy. The advanced courses may be accepted as part of the Ph.D. program of the participating institutions.

The National Science Foundation is offering grants to college teachers of introductory statistics who wish to attend the 1959 session. Stipends of $75 per week for the six weeks of the session plus $15 per week per dependent (up to four) will be made available for a maximum of 30 applicants; in addition, there will be a travel allowance of 4 cents per mile, round trip. Tuition and fees will be paid

by the National Science Foundation. Participants will normally be enrolled in classes for graduate credit.

The courses to be offered in statistics in 1959 at Raleigh are as follows: Statistical Methods I and II, Statistical Theory I, II, and III, Theory of Sampling Applied to Survey Design, Stochastic Processes and Their Applications, Methods of Operations Research, Advanced Topics in Statistical Methods.

A number of courses will, in addition, be available in the Mathematics Department. Courses to be offered during the June 8–July 17 period of the Summer Session in Statistics include: Differential Equations, Introduction to Determinants and Matrices, Introduction to Applied Mathematics, Numerical Analysis, Advanced Calculus I, Boundary Value Problems.

(For students interested in the regular 12-week summer session at North Carolina State College, additional Mathematics courses are available: Advanced Calculus II and Advanced Differential Equations will be offered in the second 6-week session beginning July 20; Complex Variable Theory and Vector Spaces and Matrices will be offered over the 12 weeks of both sessions.) The formal course offerings will be supplemented by seminars and special lectures.

Teachers under National Science Foundation grants will probably be interested primarily in the Statistical Methods and Statistical Theory sequences and in Sampling.

Applicants for National Science Foundation grants will be selected on the basis of interest in continued teaching of statistics, evidence of excellence as a teacher, previous academic record of the applicant, number of introductory statistics courses now teaching, and number of students contacted. Applications must be received not later than February 16, 1959 to be assured of full consideration. Applicants will be notified of the selection committee's action not later than March 16, 1959, and must accept (or decline) a fellowship award not later than April 1, 1959.

Requests for application blanks for the summer school and for National Science Foundation grants should be addressed to: F. E. McVay, Department of Experimental Statistics, North Carolina State College, Raleigh, North Carolina.

REPORT OF THE MONTEREY, CALIFORNIA, MEETING OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The seventy-ninth meeting of The Institute of Mathematical Statistics, a Western Region Meeting, was held in Spanagel Hall on the campus of the United States Naval Postgraduate School at Monterey, California, on November 14–15, 1958.

The Chairman of the Program Committee was Richard Link. David Stoller acted as chairman in his absence.

Sixty-one people registered for the meeting. Forty-six members of the Institute attended.

The program for the meeting was as follows:

FRIDAY, NOVEMBER 14, 1958

Welcoming Remarks, Rear Admiral E. E. Yeomans, Superintendent, U. S. Naval Postgraduate School.

10:00–12:00 a.m. Special Topics (I)

Chairman: CHARLES B. BELL, San Diego State College.

1. *Multivariate Tchebycheff Inequalities*, INGRAM OLKIN, Michigan State University.
2. *Evaluation of Extreme Tail Probabilities*, DAVID BLACKWELL AND JOSEPH HODGES, University of California.

1:30–2:30 p.m. Multiple Comparisons and Decisions

Chairman: FRED C. ANDREWS, University of Oregon.

1. *Multiple Decision-Ranking Procedures*, ROBERT E. BECHHOFER, Cornell University.
2. *Estimation of the Means of Dependent Variables*, OLIVE JEAN DUNN, Iowa State College (now at University of California, Los Angeles).

2:45–4:00 p.m. Special Topics (II)

Chairman: OLIVE JEAN DUNN, Univ. of California, Los Angeles.

1. *Statistical Problems in Radio Wave Propagation*, WILLIAM C. HOFFMAN, RAND Corporation.
2. *Stochastic Models for the Electron Multiplier Tube*, HOWARD G. TUCKER, University of California, Riverside.
3. *On a Multicompartment Migration Model with Chronic Feeding—Preliminary Report*, D. WIGGINS, Hanford Laboratory.

SATURDAY, NOVEMBER 15, 1958

10:00–12:00 a.m. Experimental Design and Analysis

Chairman: DAVID S. STOLLER, The RAND Corporation.

1. *Sequential Design of Experiments*, HERMAN CHERNOFF, Stanford Univ.
2. *Tests Associated with Poisson Processes*, BENJAMIN EPSTEIN, Wayne State University.

1:30–3:30 p.m. Contributed Papers

Chairman: HERMAN RUBIN, University of Oregon

1. *On Computing Expectations in Sequential Analysis*, FRED C. ANDREWS, University of Oregon, and J. R. Blum, Indiana University.
2. *Exact Nonparametric Tests for Randomized Blocks*, JOHN E. WALSH, Systems Development Corporation. (By title)
3. *On the Determination of Joint Distributions from the Marginal Distributions of Linear Combinations*, THOMAS S. FERGUSON, University of California, Los Angeles.
4. *Approximation to the Probability Density of Zero—Crossing Intervals of a Gaussian Process*, SYLVAIN EHRENFELD, New York University. (By title)
5. *The Probability in the Extreme Tail of a Convolution*, DAVID BLACKWELL AND J. L. HODGES, JR., Univ. of California. (By title)
6. *Asymptotic Methods of Evaluating the Integral from a to* ∞ *of* $f(x)$, WYMAN RICHARDSON, Univ. of North Carolina. (By title)
7. *Some Properties of Binary Arrays which are Generated by Iterated Sequences and Reversals*, H. VON GUERARD, Lockheed Aircraft Corporation.

GERALD J. LIEBERMAN
Associate Secretary

FINAL REPORT OF THE EDITORS OF THE ANNALS FOR 1958

The *Annals* is indebted to the following people who have generously given refereeing assistance: G. Albert, T. W. Anderson, E. Barankin, G. A. Barnard, G. Baxter, V. E. Benes, A. Birnbaum, D. Blackwell, J. Blum, R. Blumenthal, C. Blyth, L. Breiman, R. Bradt, K. A. Bush, V. J. Chacko, D. Champernowne, H. Chernoff, A. Clarke, A. C. Cohen, W. S. Connor, J. Cornfield, D. R. Cox, H. David, K. Dawson, R. Dawson, L. E. Dubins, A. Duncan, Olive J. Dunn, J. Durbin, P. S. Dwyer, G. Elfving, B. Epstein, T. Ferguson, G. H. Freeman, S. G. Ghurye, I. J. Good, B. Greenberg, U. Glenander, F. Graybill, F. Grubbs, E. J. Gumbel, S. S. Gupta, J. Hannan, Nancy Lee Hannye, M. Hansen, J. L. Hodges, Jr., W. Hoeffding, H. Hotelling, S. Hunter, G. S. James, A. Joffee, G. M. Jenkins, V. Johns, M. Juncosa, L. Katz, E. S. Keeping, D. G. Kendall, S. Kullback, R. Laha, L. LeCam, R. A. Leibler, E. Lukacs, A. Madansky, W. Madow, C. Mallows, T. K. Matthes, J. McGregor, B. McMillan, D. Mesner, A. Mood, R. B. Murphy, P. E. Ney, I. Olkin, E. Parzen, M. P. Peisakoff, K. Pillai, J. W. Pratt, W. E. Pruitt, R. Pyke, C. R. Rao, E. Reich, J. Riordan, D. Robson, Joan Rosenblatt, M. Rosenblatt, S. N. Roy, H. Rubin, J. Sacks, I. R. Savage, L. J. Savage, Rosedith Sitgreaves, J. L. Snell, M. Sobel, C. M. Stein, Charlotte Striebel, J. C. Tanner, R. F. Tate, H. Teicher, D. Teichroew, A. J. Thomasian, D. R. Truax, S. Vajda, H. R. Van der Vaart, D. Votaw, D. L. Wallace, L. Wegner, B. L. Welch, O. Wesler, R. F. White, R. Wijsman, M. B. Wilk, E. Williams, D. G. Wishart, A. Wortham, Ying Yao, M. Zelen.

December 3, 1958

T. E. Harris, *Editor*
(before July 1, 1958)
William Kruskal, *Editor*
(after June 30, 1958)

PUBLICATIONS RECEIVED

Moore, Geoffrey H., *Measuring Recessions*, Occasional Paper 61, National Bureau of Economic Research, Inc., New York, 1958 (Reprinted from the June 1958 issue of the *Journal of the American Statistical Association*), $1.00.

BIOMETRIKA

Volume 45, Parts 3 and 4 Contents December 1958

Memoirs:

BAYES, THOMAS. Essay towards solving a problem in the doctrine of chances. [Reproduced from Phil. Trans. Roy. Soc. 1763, 53, 370–418]. Studies in the history of probability and statistics. IX. With a biographical note by G. A. Barnard. LESLIE, P. H. AND GOWER, J. C. The properties of a stochastic model for two competing species. TANNER, J. C. A problem in the combination of accident frequencies. DARROCH, J. N. The multiple-recapture census. I. Estimation of a closed population. BULMER, M. G. Confidence intervals for distance in the analysis of variance. FINNEY, D. J. The efficiencies of alternative estimators for an asymptotic regression equation. PATTERSON, H. D. The use of autoregression in fitting an exponential curve. HAIGHT, FRANK A. Two queues in parallel. SHENTON, L. R. Moment estimators and maximum likelihood. SRIVASTAVA, A. B. L. Effect of non-normality on the power function of *t*-test. PEARSON, E. S. Note on Mr. Srivastava's paper on the power function of Student's test. BARTON, D. E. AND CASLEY, D. J. A quick estimate of the regression coefficient. ZINGER, A. AND ST-PIERRE, J. On the choice of the best amongst three normal populations with known variances. DRONKERS, J. J. Approximate formulae for the statistical distributions of extreme values. HOOPER, J. W. The sampling variance of correlation coefficients under assumptions of fixed and mixed variates. JOHNSON, N. L. The mean deviation, with special reference to samples from a Pearson Type III population. MERRINGTON, MAXINE AND PEARSON, E. S. An approximation to the distribution of non-central *t*. FOSTER, F. G. Upper percentage points of the generalised beta distribution. III. MENDENHALL, WILLIAM AND HADER, R. J. Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data. MENDENHALL, WILLIAM. A bibliography on life testing and related topics. *Miscellanea:* Contributions by D. E. BARTON & F. N. DAVID, D. R. COX, EDWIN L. CROW, H. E. DANIELS & M. G. KENDALL, IRWIN GUTTMAN, D. G. KABE, ROY LEIPNIK, RITA MAURICE, T. A. RAMASUBBAN, S. N. ROY & R. GNANDESIKAN, M. SANKARAN, B. V. SUKHATME

Corrigenda: N. L. JOHNSON, A. R. KAMAT, J. SAW

Reviews *Other Books received*

The subscription, *payable in advance*, is now 54s (or $8.00), per volume (including postage). Cheques should be made payable to *Biometrika* crossed "*a/c Biometrika Trust*" and sent to the Secretary, Biometrika Office Department of Statistics, University College, London, W.C.1. All foreign cheques must be drawn on a bank having a London agency.

Issued by THE BIOMETRIKA OFFICE, University College, London

ECONOMETRICA

Journal of the Econometric Society

Contents of Vol. 26, No. 4 - October 1958

HOLLIS B. CHENERY AND TSUNEHIKO WATANABE. International Comparison of the Structure of Production
KENNETH J. ARROW AND LEONID HURWICZ. On the Stability of the Competitive Equilibrium, I
L. R. KLEIN. The Estimation of Distributed Lags
IRVING HOCH. Simultaneous Equation Bias in the Context of the Cobb-Douglas Production Function
JAN TINBERGEN. L'apport de Bernard Chait à l'économétrie
HAROLD T. DAVIS. Charles Frederick Roos
MAURICE FRECHET. Letter to the Editor
REPORT OF THE PHILADELPHIA MEETING

BOOK REVIEWS:

The Accumulation of Capital (Joan Robinson). Review by L. R. Klein *International Economic Papers, No. 6* (International Economic Association). Review by Abba P. Lerner *Insurance and Economic Theory* (Irving Pfeffer). Review by Paul Boschan *Essai sur la théorie générale de la monnaie: économique rationelle* (A. Aupetit). Review by Léon Dupriez *Tables of the Non-Central t-Distribution* (*Density Function, Cumulative Distribution Function and Percentage Points*) (George J. Reskinoff and Gerald J. Lieberman). Review by N. L. Johnson *The United States Capital Position and the Structure of Its Foreign Trade* (M. A. Diab). Review by R. E. Caves *Numerical Analysis* (Zdenek Kopal). Review by H. E. Goheen *Economic Policy: Principles and Design* (J. Tinbergen). Review by Arthur Smithies *Second Symposium in Linear Programming*, Volumes I and II (H. A. Antosiewics, ed.). Review by Robert Dorfman *Structures et Cycles Economiques*, Tome II (Johan Åkerman). Review by Henri Guitton *Théories des fonctions aléatoires* (A. Blanc-Lapierre and R. Fortet). Review by L. Schmetterer *Selected Papers in Statistics and Probability by Abraham Wald*. Review by Alan Stuart *Analisis de la Demanda: Un Estudio de Econometria* (Herman Wold and Lars Jureen). Review by Angel Alcaide Inchausti *Geographic Differentials of Agricultural Wages in the United States* (Willis D. Weatherford). Review by H. Gollnick *Price, Cost, and Output* (P. J. D. Wiles). Review by J. A. Nordin *Mathematische Statistik* (van der Waerden). Review by J. Pfanzagl *Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école americaine* (M. Allais). Review by Eberhard Fels

ANNOUNCEMENTS AND NOTES

JOURNAL OF THE

ROYAL STATISTICAL SOCIETY

Series B (Methodological)

30s per part | Vol. 20 No. 2 1958 | Ann. Sub. Incl. post £3 2s 0d

CONTENTS

THE REGRESSION ANALYSIS OF BINARY SEQUENCES. By D. R. COX. (With Discussion.)
RENEWAL THEORY AND ITS RAMIFICATIONS. By WALTER L. SMITH. (With Discussion.)
ON ASYMPTOTICALLY EFFICIENT CONSISTENT ESTIMATES OF THE SPECTRAL DENSITY FUNCTION OF A STATIONARY TIME SERIES. By EMANUEL PARZEN.
THE ESTIMATION OF THE SPECTRAL DENSITY AFTER TREND REMOVAL. By E. J. HANNAN.
ON THE SMOOTHING OF PROBABILITY DENSITY FUNCTIONS. By P. WHITTLE.
EXPERIMENTS WITH MIXTURES. By HENRY SCHEFFÉ.
THE INTERACTION ALGORITHM AND PRACTICAL FOURIER ANALYSIS. By I. J. GOOD.
EQUALLY CORRELATED VARIATES AND THE MULTINORMAL INTEGRAL. By ALAN STUART.
FORMULAE FOR CALCULATING THE OPERATING CHARACTERISTIC AND THE AVERAGE SAMPLE NUMBER OF SOME SEQUENTIAL TESTS. By K. W. KEMP.
ON CORRECTIONS TO THE CHI-SQUARED DISTRIBUTION. By J. H. DARWIN.
SAMPLING WITHOUT REPLACEMENT WITH PROBABILITY PROPORTIONAL TO SIZE. By W. L. STEVENS.
MULTIVARIATE QUANTAL ANALYSIS. By P. J. CLARINGBOLD.
A CONCISE DERIVATION OF GENERAL ORTHOGONAL POLYNOMIALS. By C. P. COX.
A SIMPLIFIED MODEL FOR DELAYS IN OVERTAKING ON A TWO-LANE ROAD. By J. C. TANNER.

The Royal Statistical Society, 21, Bentinck Street, London, W. 1

METRON

Vol. XIX - N. 1-2

Contents

		Page
CORRADO GINI	*Logic in Statistics*	1
CORRADO GINI	*Gerolamo Cardano ed i fondamenti del calcolo delle probabilità*	78
A. R. KAMAT	*Contributions to the theory of statistics based on the first and second successive differences*	97
C. GINI, C. VITERBO, C. BENEDETTI, A. HERZEL	*Problemi di transvariazione inversa*	119
CARLO BENEDETTI	*Il coefficiente di correlazione del Bravais come funzione non moltiplicativa*	183
AMATO HERZEL	*Influenza del raggruppamento in classi sulla probabilità e sull'intensità di transvariazione*	199
Z. W. BIRNBAUM	*On an inequality due to S. Gatti*	243
M. DE NOVELLIS	*Some applications and developments of Gatti-Birnbaum inequality*	245

Manuscripts submitted for publication should be addressed to Prof. CORRADO GINI, Università di Roma (Italy), Via delle Terme di Diocleziano, 10. The Editors will not be responsible for the safe return of the original.